Vol. 47 NOVEMBER, 1956 No. 7

The Journal of Educational Psychology

Read

UNIVERSITY OF MICHIGAN
FEB 1
PERIODICAL READING ROOM

CONTENTS

Sex Differences in the Distribution of Teacher Approval and Disapproval Among Sixth-Grade Children 385
WILLIAM J. MEYER AND GEORGE G. THOMPSON

A Comparison of the Factorial Structures of Cognitive Functions for a High and Low Status Group 397
JAMES V. MITCHELL, JR.

Effect of Immediacy of Knowledge of Correctness of Response upon Learning 415
ARTHUR C. F. GILBERT

Effect of Teacher-Student Interviews on Classroom Achievement 424
ARTHUR J. HOEHNANDELI SALTZ

A Longitudinal Study of Sibling Resemblances in Intelligence and Achievement 436
SARAH M. SCHOONOVER

Book Reviews 443
Publications Received 446

Published monthly except June to September
$7.00 a year in U. S. and Pan America; Canada, $7.20; other countries, $7.40.
Single issues $1.20

WARWICK & YORK, INC.
BALTIMORE 2, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879.

THE JOURNAL OF Educational Psychology

Established 1910

BOARD OF EDITORS*

WM. CLARK TROW, *Chairman*
University of Michigan

STEPHEN M. COREY
Teachers College
Columbia University

J. B. STROUD
State University of Iowa

*Appointed by the American Psychological Association.

EDITORIAL CONSULTANTS

PAUL BLOMMERS
CHESTER W. HARRIS
HAROLD E. JONES
H. H. REMMERS
PERCIVAL M. SYMONDS
MILES A. TINKER
ALEXANDER G. WESMAN
PAUL A. WITTY

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY is devoted primarily to the scientific study of problems of learning, teaching, and measurement of the psychological development of the individual. THE JOURNAL will contain articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, and personality, particularly as related to school adjustment; emotion, motivation, and character; mental development and methods. This last will include tests, statistical techniques, and research techniques in cross-sectional and developmental studies.

INFORMATION FOR CONTRIBUTORS

Manuscripts and communications regarding editorial matters should be addressed to the Chairman of the Board of Editors, Wm. Clark Trow, School of Education, University of Michigan, Ann Arbor, Mich.

THE JOURNAL has set regulations regarding content and style of material published, and these should be observed in the preparation of manuscripts to be submitted.

Tables and graphs.—In keeping with the practice of most professional periodicals of like character, THE JOURNAL now requires authors to bear part of the added cost resulting from the use of tables, formulas, and graphs. Prospective contributors may obtain detailed statement as to

(Continued on Inside Back Cover)

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 47 November, 1956 Number 7

SEX DIFFERENCES IN THE DISTRIBUTION OF TEACHER APPROVAL AND DISAPPROVAL AMONG SIXTH-GRADE CHILDREN

WILLIAM J. MEYER and GEORGE G. THOMPSON

Psychology Department, Syracuse University

This study was designed to investigate the relative frequency of women teachers' approval and disapproval evaluations of sixth-grade male as contrasted with female pupils. The relevant data for this study were obtained by means of two independent techniques: thirty hours of direct observation of teacher-pupil interactions in each of three classrooms; and, the use of a modification of the "Guess Who?" technique to determine if the children themselves were aware of any sex differences in their teachers' approval and disapproval evaluations.

There is considerable agreement among psychologists that the use of approval by the teacher results in better learning and probably in better over-all adjustment (*14*, *18*). Some studies (*19*, *24*) have shown that personal maladjustments in teachers have deleterious effects on the adjustment level of the children in their classes.

In a series of studies by Anderson *et al.* (*1–3*), using a sample of kindergarten-age children, the data indicate that teachers typically use statements of a dominative nature in their interactions with the children in their classes. Anderson further reports that the teachers in his study tended to levy most of their dominative and/or integrative overtures on only a few pupils to the relative neglect of the other children in the classroom. Further evidence of this nature is reported in a study by deGroat and Thompson using four sixth-grade classrooms (*10*). In addition to reporting inequities in teacher approval and disapproval they also found that teachers give more praise to the youngsters who are brighter, better adjusted and higher achievers. The more poorly adjusted

and the duller children were observed by these investigators to receive more disapproval from their teachers.

The purpose of the present investigation is to shed more light on the ways in which teachers respond toward the pupils in their classrooms. Extensive research findings have been reported in the literature (*11*, *12*, *20*, *21*, *28*) which consistently show that boys are more aggressive and generally more "unmanageable" than girls. It is our hypothesis that this "masculine" behavior will result in male pupils receiving a larger number of dominative, or punitive, contacts than girls from their teacher, who is usually a woman from the middle socio-economic stratum of our society. That is, we feel that the behavior of boys in the typical classroom is of such a nature as to make it less acceptable to teachers who probably attempt to perpetuate certain middle-class standards of what "good" classroom behavior should be. We believe that girls usually display behavior more in conformity to the standards perceived as "good" by the average elementary school teacher and will therefore receive fewer disapproval contacts and more approval contacts from their teachers.

Assuming that the above hypotheses are supported by the data we would also predict that children of elementary school age will recognize, and take for granted, that boys receive more disapproval and blame from their teachers than girls.

EXPERIMENTAL PROCEDURE

In order to test the hypothesis that boys receive a larger number of dominative, or disapproval, evaluations from their teachers than do girls, teacher-pupil interaction within three sixth-grade classrooms were recorded for a total sample of thirty hours per classroom. These time samples of classroom behavior were spread over an entire school year. Among other things being studied, interactions between teachers and pupils were classified into two categories: (a) praise contacts (teacher initiated interactions with a child in which she verbally expressed approval of some behavior which the child had displayed), and (b) blame contacts (teacher initiated interactions with a child in which she verbally expressed disapproval for some bit of behavior which the child had displayed). Observer agreement for the praise classification ranged from eighty-four to one hundred per cent with a median of approximately ninety-two per cent. Observer agreement for the blame classifica-

tion ranged from fifty-seven to one hundred per cent, with a median of approximately ninety-three per cent.

In an attempt to cast some light on children's perceptions of any sex differences in teacher disapproval, a modified "Guess Who?" approach was employed. The "Guess Who?" approach used in this study required each child to nominate fellow class members for a number of situations in which children are receiving approval or disapproval from their teacher for some behavior. (See deGroat and Thompson (*10*) for a more complete description of these scales and information about their reliabilities.) The behavior descriptions were selected on the basis of their familiarity to children and contain a fairly representative sample of situations in which children typically receive either approval or disapproval from their teachers. Each child was required to list the names of four of his classmates whom he thought fitted each of the behavior descriptions most adequately.

RESULTS

Fisher's t test was used to determine the reliability of the obtained sex differences.[1] The difference between disapproval contacts received by boys and by girls from their teachers was statistically significant in each of the three classrooms. As predicted, the boys received the larger number of disapproval contacts. These differences may be interpreted according to our hypothesis as supporting the notion that teachers are responding with counter-aggression to the greater expression of aggression by boys. The results obtained in analyzing the teachers' praise contacts with boys and girls are presented in Table II. The only statistically significant differences obtained for this variable was in school B. However the boys received more praise than the girls in each of the classrooms. It may be that the teachers are attempting to reinforce any positive behavior that the boys may display. Or this tendency to praise boys more than girls may reflect compensatory behavior for guilt feelings created in the teacher by her excessive aggressive-

[1]This test assumes that the samples being compared are homogeneous with respect to their variances. Frequently this assumption had to be rejected in some of the group comparisons. In such cases a more conservative test of significance was used which makes some allowance in the error term for heterogeneity of variance. This technique is presented in detail in Cochran and Cox (*6*).

TABLE I—SEX DIFFERENCES IN FREQUENCY OF TEACHERS' DISAPPROVAL CONTACTS

	Classroom A		Classroom B		Classroom C	
	Boys	Girls	Boys	Girls	Boys	Girls
N	10	9	12	14	17	16
Mean	11.10	2.67	10.75	2.79	10.06	1.44
S.D.	7.62	2.18	9.27	2.42	14.42	1.59
t	3.20**		3.11**		2.37*	
F	12.22**		14.67**		130.36**	
$t_{.01}$†	3.30		3.10		2.92	

* Significant at the five per cent level.
** Significant at the one per cent level.
† See footnote 2.

TABLE II—SEX DIFFERENCES IN FREQUENCY OF TEACHERS' APPROVAL CONTACTS

	Classroom A		Classroom B		Classroom C	
	Boys	Girls	Boys	Girls	Boys	Girls
N	10	9	12	14	17	16
Mean	9.90	9.67	10.50	5.50	3.71	2.69
S.D.	6.10	7.65	5.33	2.53	2.95	1.99
t	0.074		2.50*		1.15	
F	1.57		4.43**		2.19	
$t_{.05}$†	—		2.19		—	

* Significant at the five per cent level.
** Significant at the one per cent level.
† See footnote 2.

ness towards boys. Either interpretation, or any one of the several others that could be offered, is highly speculative.

The data presented above are based on the extensive observations of an objective observer who played no functional rôle in the classrooms. The data presented in the following section reflect the teachers' approval and disapproval contacts as viewed by their pupils.

"GUESS WHO?" DATA

Analysis of the "Guess Who?" data was performed along the same lines as the data obtained by direct observation. A comparison

TABLE III—SEX DIFFERENCES IN CHILDREN'S NOMINATIONS FOR TEACHER DISAPPROVAL

	Classroom A		Classroom B		Classroom C	
	Boys	Girls	Boys	Girls	Boys	Girls
N	10	9	12	14	17	16
Mean	21.60	5.33	42.33	9.71	33.82	5.18
S.D.	13.33	5.68	43.06	8.13	55.87	8.26
t	3.39**		2.79**		2.03	
F	5.51**		28.07**		19.21**	
$t_{.01}$†	3.26		3.10		—	

** Significant at the one per cent level.
† See footnote 2.

of the pupils' nominations of their peers on the disapproval items revealed statistically significant differences between boys and girls for two of the three schools. This can be interpreted as showing that the boys are viewed by the girls as well as by their male peers as being involved in more situations which evoke disapproval from their teachers.

Analysis of the children's responses to the items related to teacher approval produced no significant differences between boys and girls.

A final analysis of the "Guess Who?" data was performed in an attempt to determine how boys as contrasted with girls perceived the teacher's approval and disapproval biases. The choices

TABLE IV—SEX DIFFERENCES IN CHILDREN'S NOMINATIONS FOR TEACHER APPROVAL

	Classroom A		Classroom B		Classroom C	
	Boys	Girls	Boys	Girls	Boys	Girls
N	10	9	12	14	17	16
Mean	11.00	34.33	23.58	31.42	23.71	21.60
S.D.	12.93	32.64	18.51	25.08	35.75	25.70
t	2.09		0.856		0.242	
F	6.376**		1.835		19.35**	
$t_{.01}$†	—		—		—	

** Significant at the one per cent level.
† See footnote 2.

TABLE V—CHOICES MADE BY BOYS AND BY GIRLS ON TEACHER DISAPPROVAL ITEMS

	Classroom A				Classroom B				Classroom C			
	Boys Choosing		Girls Choosing		Boys Choosing		Girls Choosing		Boys Choosing		Girls Choosing	
	B	G	B	G	B	G	B	G	B	G	B	G
N	8	9	8	9	14	12	14	12	17	16	17	16
%	89.77	23.23	83.09	16.91	88.37	11.63	73.05	26.95	88.70	11.20	82.21	17.79
CR	5.60**		3.64**		6.10**		2.65**		7.03**		4.84**	

** Significant at the one per cent level.

made by the boys and by the girls for the approval and disapproval items were separately analyzed. It seemed unreasonable to use the t test in this situation because of the unequal numbers of boys and girls in the classroom. Therefore the groups were equated by converting the frequencies of nominations to percentages and working with percentage differences.

The results of the statistical analysis of boys' nominations on the disapproval items show that boys respond as if they usually received more blame from teachers than do girls. It would appear that boys are quite sensitive to the disapproval of their teachers. Table V shows that the girls also respond as if boys receive more teacher disapproval.

There is little consistency in the nominations made by the boys for the praise items. In schools B and C the boys react as though

TABLE VI—CHOICES MADE BY BOYS AND BY GIRLS ON TEACHER APPROVAL ITEMS

	Classroom A				Classroom B				Classroom C			
	Boys Choosing		Girls Choosing		Boys Choosing		Girls Choosing		Boys Choosing		Girls Choosing	
	B	G	B	G	B	G	B	G	B	G	B	G
N	8	9	8	9	14	12	14	12	17	16	17	16
%	36.98	63.01	34.07	65.92	56.95	43.04	24.58	75.42	57.56	42.43	46.22	53.77
CR	1.11		1.38		0.72		3.01**		0.88		0.44	

** Significant at the one per cent level.

they typically receive more praise than girls, although this difference is not statistically significant. In contrast to the boys' responses, the girls feel that they receive more praise, particularly in school B where the difference is statistically significant. These results might be interpreted as meaning that children fail to recognize any definite dichotomy in the teacher's distribution of praise contacts.

DISCUSSION

The general findings of this study support the hypothesis that the male pupil receives reliably more blame from his teacher than the female pupil. Moreover, the boys recognize that they are the recipients of a higher incidence of teacher disapproval. We feel that these data lend indirect support to the notion that "masculine" behavior is not tolerated by the typical teacher who in turn attempts to inhibit such behavior by means of punishment.

Davis and Havighurst (*8*) have discussed at length the divergence of cultural mores between lower-class children and their middle-class teachers. Their work may best be summarized in the assertion that the goals defined by the middle-class teacher do not receive reinforcement from the lower-class child's peer group or from his family. Teacher initiation of punishment for "misbehavior" only serves to reinforce an already existing dislike for school and further leads to peer group reinforcement. A similar (but by no means identical) interpretation appears relevant to the present discussion. Our society's definitions of acceptable male and female behavior are divergent particularly with respect to aggression. For example Radke (*20*) in her monograph on the relationship of parental authority to child behavior reports that the fathers in her sample felt that aggressive, assertive behavior on the part of boys was less undesirable than the identical behavior in girls (and in many cases was deemed highly desirable). The mothers felt that aggression was unacceptable behavior in either sex but in general they were in agreement that aggressive behavior is more unacceptable in girls. In another study specifically related to the notion that aggressive behavior is more unacceptable in the female culture is a study by Sears *et al.* (*22*). These writers predicted that in father-absent homes, wherein the child is brought up by the mother, boys would be less aggressive than in father-present homes in which the boy models his behavior after the father. The results

of their study support the "sex-typing" hypothesis as presented above. Bach (*4*) reports similar evidence in support of the "sex-typing" hypothesis.[2] Apparently the social mores of the typical female teacher, at least with respect to aggressive, assertive behavior, are in sharp contrast to the behavioral tendencies of the typical male youngsters. The behavioral tendencies of the female child are, however, in close agreement with those of her teachers. We feel that the above generalization accounts to a high degree for the data reported in this study. Our argument becomes somewhat stronger when the work of Wickman (*29*) and a follow-up study by Mitchell (*17*) are included in the discussion. These investigators found that teachers perceive aggressive nonconforming behavior as more serious than withdrawal behavior. More recently Kaplan (*15*) has reported that the aggressive child was deemed annoying to almost three-quarters of the teachers in his sample. The present investigation suggests that perhaps teachers react to the aggressive behavior of children with counter-aggression, a vicious circle for both pupil and teacher.

Consistent with the above interpretation is the larger amount of variation found among the male pupils as contrasted with the female pupils. In a culture such as ours in which the father is away from the home during most of the child's waking hours (and in some instances pays only cursory attention to the youngster when at home), it appears obvious that both the male and female child are more directly influenced by the mother. Many boys, however, will be influenced more by their fathers and peer culture than by the mother because of identification with the masculine rôle in our culture. Our belief is that these more "masculine" boys are the ones who receive the greater share of teacher disapproval. Such an interpretation appears consistent with the work of Sears (*22*) and Bach (*4*).

The foregoing discussion has certain implications for the student of child development and education. If our interpretation of the teacher and male-pupil relationship is accurate, then the fact that boys dislike school more than girls is understandable. The daily punishment received by the boy for behavior he really does not consider "bad" must certainly be anxiety producing. If the

[2] Though there is insufficient evidence at this time the present writers feel that the factor of innate sex differences in aggressive tendencies should not be overlooked. See Beach (*5*) for suggestive findings.

anxiety created in the school situation becomes sufficiently intense, it seems reasonable that tension reduction can be achieved by means of avoiding school. It is known that more boys leave school at an earlier age than girls (*26*).

Perhaps of even more importance is the effect of this teacher-disapproval generated anxiety on the general personality adjustment of male pupils. It is unfortunate that we do not have evidence on the changes in adjustment level of the children in our sample, but studies by Ojemann and Wilkinson (*18*) and others indicate that consistent teacher dominance has deleterious effects on the adjustment of children. We can only speculate as to the nature of these adjustment problems but such behavioral manifestations as nervousness, withdrawal and lack of self-confidence are a few of the known symptoms.

We feel that the consistent trends in our findings imply that teachers' negative attitudes towards their male pupils arise from a lack of appreciation for the term "normal" male child. In our culture, aggressive outgoing behavior is as normal in the male as quiescent nonassertive behavior is in the female. The teacher who attempts to thwart this behavior by means of threats and punishment can only meet with frustration since the boy is confronted with a conflicting social code. A more reasonable plan to follow would seem to be one in which the excess energy and tensions of the male child could be discharged on some constructive activity. Planned physical education classes will do much to dissipate aggressive needs in a socially acceptable manner. Perhaps most important of all, however, is the knowledge that some degree of aggressive behavior is a normal part of development in both boys and girls and should be treated not as a personal threat to the teacher but as sign of "normal" social and personality development.

SUMMARY

The purpose of this study was to investigate sex differences in teacher distribution of approval and disapproval among three sixth-grade classrooms. Data relevant to the children's perceptions of their teachers' attitudes towards boys and girls were also collected. Using the discrepancies in attitude between males and females in our culture toward aggressive behavior as the basic underlying variable, the hypothesis was offered that boys, who

are more aggressive and nonconforming than girls, would receive more disapproval contacts from their teachers than girls. Girls being quiescent and more conforming than boys would as a consequence receive more approval from their teachers than boys. We further hypothesized that both boys and girls will be aware of the differences in their teachers' attitudes towards them.

In order to test the foregoing hypothesis three sixth-grade teachers and their pupils were directly observed for a total of thirty hours per classroom. All teacher initiated contacts of an approval or disapproval nature were recorded. The measurement of the children's perceptions of teacher attitude was accomplished by means of a variation of the "Guess Who?" technique. The pupils were asked to list the names of four students who best fitted a series of statements of a teacher approval nature and of disapproval nature. Analysis was made of the number of children of each sex chosen for the approval items and for the disapproval items.

Statistical analysis of the data clearly supports our hypothesis with respect to male pupils. In all three schools the boys received reliably more disapproval from their teachers than the girls. We also found that both the boys and the girls nominated more boys for the disapproval items than girls. This difference was statistically reliable. With respect to the second hypothesis concerning girls, the data did not yield any clear-cut differences. If any trend was present it was in a direction opposite to that predicted. These results indicated that the teachers in our sample tended to have fewer contacts with the girls in their classrooms.

The results of this investigation were interpreted as being consistent with the notion of a sex difference in attitude towards aggressive behavior. The conclusion was drawn that teachers attempt to "socialize" the male child by means of dominative counter-aggressive behavior. The negative consequences of this situation for the child are discussed.

REFERENCES

1) H. H. Anderson and H. M. Brewer, "Dominative and Socially Integrative Behavior of Kindergarten Teachers," *Applied Psychology Monograph No. 6*. Stanford, Calif., Stanford University Press, 1945.

2) H. H. Anderson and J. E. Brewer, "Effects of Teachers' Dominative and Integrative Contacts on Children's Classroom Behavior," *Applied*

Psychology Monograph No. 8. Stanford, Calif., Stanford University Press, 1946.

3) H. H. Anderson, J. E. Brewer, and M. F. Reed, "Studies of Teachers' Classroom Personalities: III. Follow-up studies of the effects of dominative and integrative contacts on children's behavior," *Applied Psychology Monograph No. 11.* Stanford, Calif., Stanford University Press, 1946.

4) G. R. Bach, "Father-Fantasies and Father-Typing in Father Separated Children," *Child Development*, 17: 63–80, 1946.

5) F. A. Beach, *Hormones and Behavior.* New York, Paul B. Hoeber, 1948.

6) W. G. Cochran and G. M. Cox, *Experimental Designs.* New York, Wiley & Sons, Inc., 1950.

7) W. A. Davis, *Social-Class Influences upon Learning.* Cambridge, Harvard University Press, 1948.

8) W. A. Davis and R. J. Havighurst, *Father of the Man: How Your Child Gets His Personality.* Boston, Houghton Mifflin Co., 1947.

9) W. A. Davis and J. Dollard, *Children of Bondage.* Washington, D. C., American Council on Education, 1940.

10) A. F. deGroat and G. G. Thompson, "A Study of the Distribution of Teacher Approval and Disapproval among Sixth-Grade Pupils," *Journal of Experimental Education*, 18: 57–75, 1949.

11) F. L. Goodenough, "Anger in Young Children," *Child Welfare Monograph No. 9*, pp. xiii + 278. Minneapolis, University of Minnesota Press, 1931.

12) M. L. Hayes, "A Study of the Classroom Disturbances of Eighth Grade Boys and Girls, *Teachers College Contributions to Education*, 871: 139, 1943.

13) E. Hollenberg and M. Sperry, "Some Antecedents of Aggression and Effects of Frustration on Doll Play," *Personality*, 1: 32–43, 1951.

14) E. Hurlock, "The Value of Praise and Reproof as Incentives for Children," *Archives of Psychology*, 11: No. 71, 1924.

15) L. Kaplan, "The Annoyances of Elementary School Teachers," *Journal of Educational Research*, 45: 649–665, 1951–52.

16) C. Leeds and W. Cook, "The Construction and Differential Value of a Scale for Determining Teacher-Pupil Attitudes," *Journal Experimental Education*, 16: 149–159, 1947.

17) J. C. Mitchell, "A Study of Teachers' and of Mental-Hygienists' Ratings of Certain Behavior Problems of Children," *Journal of Educational Research*, 36: 292–307, 1942.

18) R. Ojemann and F. R. Wilkinson, "The Effect on Pupil Growth of an Increase in Teacher's Understanding of Pupil Behavior," *Journal of Experimental Education*, 8: 143–147, 1939.

19) W. Olson and M. Wilkinson, "Teacher Personality as Revealed by the Amount and Kind of Verbal Direction Used in Behavior Control," *Educational Administration and Supervision*, 24: 81–93, 1938.

20) M. J. Radke, "The Relation of Parental Authority to Children's Behavior and Attitudes," *Child Welfare Monograph No. 22*, Minneapolis, University of Minnesota Press, 1946.

21) P. S. Sears, "Doll Play Aggression in Normal Young Children: Influence of Sex, Age, Sibling Status, Father's Absence," *Psychological Monographs*, 65: 42, 1951.

22) R. R. Sears, M. H. Pintler, and P. Sears, "Effect of Father Separation on Preschool Children's Doll Play Aggression," *Child Development*, 17: 219–243, 1946.

23) R. Sears, J. Whiting, V. Nowlis, and P. Sears, "Some Child-rearing Antecedents of Aggression and Dependency in Young Children," *Genetic Psychology Monographs*, 47: 135–234, 1953.

24) W. Snyder, "Do Teachers Cause Maladjustment?," *Journal Exceptional Children*, 14: 40–46, 1947.

25) R. T. Sollenberger, "Some Relationships between the Urinary Excretion of Male Hormone by Maturing Boys and Their Expressed Interests and Attitudes," *Journal of Psychology*, 9: 179–189, 1940.

26) S. Tenenbaum, "Uncontrolled Expressions of Children's Attitudes toward School," *Elementary School Journal*, 40: 670–768, 1939–40.

27) S. Tiedeman, "A Study of Pupil-Teacher Relationship," *Journal of Educational Research*, 35: 657–664, 1942.

28) R. D. Tuddenham, "Studies in Reputation: I. Sex and Grade Differences," *Psychology Monographs*, 66: 58, 1952.

29) E. K. Wickman, *Teachers and Behavior Problems*. New York, Commonwealth Fund, 1938.

30) J. Withall, "The Development of a Technique for the Measurement of Social-Emotional Climate in Classrooms," *Journal of Experimental Education*, 17: 347–361, 1949.

A COMPARISON OF THE FACTORIAL STRUCTURE OF COGNITIVE FUNCTIONS FOR A HIGH AND LOW STATUS GROUP[1]

JAMES V. MITCHELL, JR.

Miami University

Comparative studies dealing with the similarities and differences between the structure of mental abilities for different groups may be of value in several ways. For those who are interested in the efficacy and consistency of a particular method of factor analysis, such comparisons become tests of the "stability" or invariance of a factorial solution for groups of dissimilar character (*8*, *10*, *11*). For those whose primary interest is in the nature, development, and organization of psychological traits, such comparisons are a means of determining some of the crucial variables associated with the development and organization of particular kinds of trait patterns (*1*). For those who are especially interested in group differences, such comparisons permit a more comprehensive and analytical approach to the patterning and interrelationships among significant variables (*2*, *5*, *12*). And for those who are concerned about the appropriateness of certain types of tests for certain kinds of populations, such comparisons furnish data which can be of invaluable assistance in making the correct choice of tests. The findings of the present study may be interpreted in the light of any one of these considerations.

The purpose of this study was to compare the factorial organization of mental abilities for a well-defined High and Low Status Group. This necessitated the selection of a test battery that adequately represented the domain of mental abilities as charted by previous factorial experiments, the administration of this battery to a carefully selected High Status and Low Status Group, the computation of intercorrelation coefficients among the tests for

[1] This study is based on a Ph. D. dissertation done under the direction of Dr. Allison Davis of the University of Chicago.

each of the two groups, and the factorization of both correlation matrices with subsequent rotation of reference axes. The factor matrices so obtained were to be subject to a comparative analysis which would consider all evidence of significant similarities or differences in factor patterns but which would give special consideration to the following questions:

(1) What proportion of the common factor variance for both groups is accounted for by a general first-order factor corresponding to Spearman's general intellective factor "g"?

(2) Are essentially the same factors identified in the analyses for the two status groups?

(3) Is there any difference in the extent to which the "simple structure" criterion can be attained in applying rotational procedures to the two groups?

(4) What comparisons can be made between the organization of mental abilities for each group as determined by the correlations among the separate factors or abilities?

(5) Are the factorial compositions of individual tests in the battery approximately the same for the two groups, or are there differences that suggest that some of the tests represent different functions or organizations of such functions for the two groups?

PROCEDURE

The subjects of this study were eleven- and twelve-year-old pupils in certain selected schools located in and around the city of Chicago. Only those schools were selected whose student populations seemed likely to contain a large proportion of either "Low" or "High" status people as defined by the index of status employed in this study. All of the eleven- and twelve-year-old pupils in these selected schools were administered a battery of mental tests consisting of the following well-known and widely used instruments: the Addition, Multiplication, Vocabulary, Completion, Figures, Cards, First Letters, Four-Letter Words, Suffixes, Letter Series, Letter Grouping, and Pedigrees tests of the *Chicago Tests of Primary Mental Abilities* for ages eleven to seventeen; the Spatial Relationships, Logical Reasoning, Numerical Reasoning, and Verbal Concepts subtests which comprise the elementary form of the *California Short-Form Test of Mental Maturity*, but which were treated as separate tests for purposes of analysis; the *Otis Quick-Scoring Mental Ability Test* (Beta, Form Dm); and the *Henmon-Nelson Test of Mental Ability* (Elementary, Form A). The testing was done in four sessions to avoid the effects of fatigue.

The tests from the Chicago series are instruments of demonstrated factorial purity, and they were included in the present battery so that any evidence of differences in mental organization for the two status groups would be clearly revealed in the resulting configurations. The four subtests of the California test, which are also supposed to be representative of separate abilities or factors, were included for essentially the same reason, and it was judged that these sixteen tests would bear the major burden of defining and allocating the common-factor space for each of the status groups. The Otis Beta and Henmon-Nelson tests were included in the battery so that any differences in the rôle of the so-called "IQ" test in the mental organization of the separate status groups would be clearly revealed in the analysis.

At the end of the first testing session each subject received a questionnaire which he was requested to take home and have his parents complete. The most important function of this questionnaire was to determine the father's occupation and the address of the family dwelling, both of which were necessary to determine the index of status that was employed in this study. This index was a modified form of the "Index of Status Characteristics" developed by Warner, Meeker and Eells, which included scales for rating the father's occupation, the house type, the dwelling area and the source of income (*15*). The latter variable was omitted from the index used for the present study, because of anticipated resistance to questions concerning source of income, and the three remaining variables were combined to form a modified Index in accordance with a method suggested by Warner *et al.* for computing the Index when this variable is unavailable. The father's occupation was obtained directly from the questionnaires, and the type of house lived in and the nature of the dwelling area were easily ascertained by actually visiting the location after obtaining the address from the questionnaire and then observing and rating while on the scene. Since each child had an "ISC," or Index of Status Characteristics, which indicated his relative position on the status continuum, "Low" and "High" status groups could be defined by establishing certain critical ISC scores as limits for each group. Accordingly, the High Status Group was designated as occupying that part of the status continuum extending from an Index value of 33 or less, while the Low Status Group was designated as occupying that part of the continuum extending

from an Index value of 54 or more. (Low Index values are indicative of high status.) For the original population on which the ISC was validated, these critical values designated subjects who were upper middle class or above in the one case and upper lower class or below in the other (*15*). The ISC averages for the Low and High Status groups were 67.6 and 22.1 respectively, and the status range extended from upper-middle status to upper-upper status for the High Status Group and from lower-lower status to upper-lower status for the Low Status Group. There were one hundred and seventy-nine subjects in the High Status Group and one hundred and seventy-two subjects in the Low Status Group thus defined. "Status," as employed herein and measured by the ISC, actually refers to socio-economic status, but the correlation between socio-economic level and social status is high enough (0.966) to justify using these terms interchangeably for our present purposes (*15*).

In preparing the data for analysis, coefficients of correlation were computed for all possible test pairs within each of the status groups, and these correlations were recorded in the 18 x 18 correlation matrices that were prepared separately for both the Low and High Status Groups.[2] Both of these matrices were factored by means of the centroid method of multiple factor analysis as described by L. L. Thurstone, and the factor matrices so obtained were subject to rotational procedures until oblique solutions were achieved which best satisfied the criteria for simple structure (*14*). The resulting factor patterns for the two groups were then compared.

Because the correlations between primary factors are determined by the placement of the reference axes in accordance with simple structure criteria, which is at best a somewhat inexact procedure, those correlations tend to be rather unstable. Therefore, in order to obtain reliable data bearing on the correlational relationships between the factors that were extracted, multiple group analyses of the correlation matrices for both status groups were undertaken, employing for this purpose the factorial clusters that had been identified in the centroid analysis (*14*). This proce-

---

[2] The original correlation matrices have been deposited with the American Documentation Institute. Order Document No. 4866 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit $1.25 for 35 mm microfilm or $1.25 for photoprints.

dure yielded correlations between factors that were unique and mathematically determinate and which could be used confidently for purposes of comparison.

RESULTS AND INTERPRETATION

The centroid factor matrices that resulted from the factorization of the correlation matrices for the two status groups are shown in Table I (for the Low Status Group) and Table II (for the High Status Group). Factor extraction was regarded as complete when two-thirds of the loadings for a factor fell below twice their standard errors (as determined by Burt's empirical formula) (*13*), and when the product of the two highest loadings for a factor fell below $\frac{1}{\sqrt{N}}$ (*13*). Application of these combined criteria indicated that factorization should be concluded for both groups with the extraction of the sixth factor, but a seventh factor was also extracted to make absolutely certain that further meaningful factors

TABLE I—CENTROID FACTOR MATRIX: LOW STATUS GROUP

Test	Factor							
	I	II	III	IV	V	VI	VII	h^2
PMA Addition	59	28	−32	−24	28	13	−15	70
PMA Multiplication	63	32	−27	−22	24	14	−11	71
PMA Vocabulary	76	15	17	16	−03	−20	−15	72
PMA Completion	78	03	17	−06	−07	−15	04	67
PMA Figures	48	−42	−27	21	−19	05	−15	58
PMA Cards	58	−44	−37	19	−28	03	05	78
PMA First Letters	74	29	10	24	−05	23	11	77
PMA Four-Letter Words	59	37	−13	−09	−28	05	16	62
PMA Suffixes	61	29	17	07	−30	24	−10	65
PMA Letter Series	64	−20	12	23	40	07	19	72
PMA Letter Grouping	53	09	−14	24	32	−11	23	53
PMA Pedigrees	66	−04	10	26	16	−04	−07	55
Calif. Spatial Relations	56	−19	−14	−03	−04	−05	12	39
Calif. Logical Reasoning	71	−22	09	−19	−01	−04	−07	60
Calif. Numerical Reasoning	74	−15	05	−20	13	11	09	65
Calif. Verbal Concepts	84	08	28	−04	−06	−08	05	80
Otis Beta	89	−09	24	−24	−02	03	03	92
Henmon-Nelson	87	−15	20	−21	−07	16	10	90
Per cent of Total Variance	47	06	04	04	04	02	02	

TABLE II—CENTROID FACTOR MATRIX: HIGH STATUS GROUP

Test	Factor							
	I	II	III	IV	V	VI	VII	h^2
PMA Addition	61	32	−29	−22	33	12	11	74
PMA Multiplication	56	40	−29	−23	37	22	13	81
PMA Vocabulary	73	22	45	17	−10	07	10	84
PMA Completion	76	07	27	30	−04	−11	15	78
PMA Figures	44	−26	−50	28	−22	−03	23	69
PMA Cards	52	−43	−41	25	−19	10	16	76
PMA First Letters	50	42	−15	−04	−23	−12	−07	52
PMA Four-Letter Words	42	30	−20	−12	−27	−24	−16	48
PMA Suffixes	55	33	−14	03	−13	12	−30	55
PMA Letter Series	64	−20	15	−28	−24	13	−10	64
PMA Letter Grouping	49	−16	06	−37	−15	06	09	44
PMA Pedigrees	74	−18	25	−21	−05	11	−09	71
Calif. Spatial Relations	56	−27	−12	14	14	07	−21	49
Calif. Logical Reasoning	54	−33	18	06	24	−13	−15	53
Calif. Numerical Reasoning	70	−29	−04	−04	21	−14	08	65
Calif. Verbal Concepts	82	20	35	16	05	09	11	88
Otis Beta	86	−09	22	10	13	−10	04	83
Henmon-Nelson	80	−05	24	12	13	−06	−11	75
Per cent of Total Variance	41	08	07	04	04	02	02	

could not be identified. After the rotation of reference axes, it was found that the seventh factor was a residual factor of no importance, and it was therefore discarded for both groups. The first six factors represented sixty-seven per cent of the total test variance for the Low Status Group and sixty-six per cent of the total test variance for the High Status Group.

Inspection of Tables I and II reveals an interesting difference in first factor variance for the two groups. This first unrotated centroid factor, which roughly corresponds to Spearman's general intellective factor "g," accounts for some forty-seven per cent of the total test variance for the Low Status Group and only forty-one per cent of the total test variance for the High Status Group. This amounts to about seventy per cent of the common-factor variance for the Low Status Group and sixty-two per cent of the common-factor variance for the High Status Group. This difference of six per cent for the total variance and eight per cent for the common-factor variance indicates that for the Low Status

TABLE III—FINAL FACTOR MATRIX: OBLIQUE SOLUTION FOR LOW STATUS GROUP

Test	Factor					
	V	WF	S	R	N	GT
PMA Addition	00	−04	01	00	52	00
PMA Multiplication	00	01	−01	00	48	01
PMA Vocabulary	40	02	06	05	11	−14
PMA Completion	23	−06	−01	−07	−05	17
PMA Figures	06	01	62	03	09	−07
PMA Cards	−04	−01	66	−07	−05	04
PMA First Letters	−07	42	−02	24	02	−07
PMA Four-Letter Words	00	18	04	−25	−03	10
PMA Suffixes	07	43	06	00	05	−02
PMA Letter Series	−12	03	−06	50	−01	06
PMA Letter Grouping	−02	−05	−03	27	06	−11
PMA Pedigrees	17	06	08	29	11	−13
Calif. Spatial Relations	00	−11	21	−05	−04	17
Calif. Logical Reasoning	15	−15	12	−05	04	29
Calif. Numerical Reasoning	−07	−06	00	08	04	35
Calif. Verbal Concepts	21	04	−09	01	−07	19
Otis Beta	10	−04	−04	−03	−03	39
Henmon-Nelson	−06	06	00	01	−09	44

Group there is more of a communality of function between the various tests that make up the battery, and, if the present test battery adequately represents the domain of mental abilities, it is strong evidence for a less differentiated organization of mental abilities for the Low Status Group. Apparently the Low Status child approaches his various mental activities with a relatively unspecialized intellectual structure, and his "general" intelligence (or "g") is in fact a relatively more generalized kind of function than its High Status counterpart and plays a correspondingly greater rôle in the solution of all of the more specialized types of intellectual problems as well as those of a general nature. These observations are also confirmed by other data that will be presented.

Rotation of reference axes in accordance with simple structure criteria resulted in the oblique solutions shown in Table III for the Low Status Group and Table IV for the High Status Group. The oblique solution was considered preferable for both groups, since inspection of the correlation matrices disclosed that the

TABLE IV—FINAL FACTOR MATRIX: OBLIQUE SOLUTION FOR HIGH STATUS GROUP

Test	Factor					
	V	WF	S	R	N	GT
PMA Addition	−02	03	02	00	61	07
PMA Multiplication	01	−05	−01	00	69	−01
PMA Vocabulary	62	00	−05	08	−03	−09
PMA Completion	55	10	12	−12	−05	12
PMA Figures	01	07	68	−08	03	02
PMA Cards	−02	−09	64	05	−03	−01
PMA First Letters	11	36	02	−01	08	−06
PMA Four-Letter Words	−07	46	−01	01	−02	05
PMA Suffixes	05	07	−10	03	00	−15
PMA Letter Series	−03	02	01	52	−08	−06
PMA Letter Grouping	−07	08	09	47	11	00
PMA Pedigrees	08	−04	−08	40	−02	06
Calif. Spatial Relations	−07	−14	06	−02	−07	18
Calif. Logical Reasoning	02	−03	−11	−02	−11	40
Calif. Numerical Reasoning	00	07	14	03	14	39
Calif. Verbal Concepts	57	−04	−03	03	10	−01
Otis Beta	31	06	02	00	03	28
Henmon-Nelson	27	02	−10	−01	−05	24

inter-factor correlations would probably be too high to justify the imposition of orthogonal structure on the test configurations. Essentially the same factors were identified for both groups, and these factors, with but one exception, represented the primary mental abilities that have been reported by Thurstone and others: The "V" or Verbal Meaning factor, which represents a knowledge of words and their meanings, had its highest loadings for the Vocabulary and Completion tests of the Chicago PMA series and the Verbal Concepts test of the California for both status groups. The "WF" or Word Fluency factor, which represents the ability to call quickly to mind the word forms which are appropriate or required for given circumstances or conditions, had its highest loadings for the First Letters, Four-Letter Words, and Suffixes tests for the Low Status Group, while the same factor was determined primarily by the First Letters and Four-Letter Words tests for the High Status Group. The "S" or Space factor, which involves the comprehension of spatial relations or the ability to visualize the manipulation of objects in space, was determined

primarily by the Figures and Cards tests of the Chicago PMA series for both status groups. The "R" or Reasoning factor, which refers to the skill required to discover a principle or relationship from a given set of data and then to make a further application of it, was determined primarily by the Letter Series, Letter Grouping, and Pedigrees tests of the Chicago PMA series for both status groups. The "N" or Number factor, which represents the ability to perform the simple numerical operations in addition, subtraction, etc., had its highest loadings for the Addition and Multiplication tests of the Chicago PMA series for both status groups. Also extracted from the correlation matrices for both status groups was a factor that had its highest loadings for the Otis Beta and Henmon-Nelson tests and the subtests for the California Test of Mental Maturity. Since the Otis Beta and Henmon-Nelson tests are both tests of general intelligence characterized by a high degree of factorial complexity, it is probably best to designate this factor as the "GT" or General Test factor. It is noteworthy that the subtests of the California, which are supposed to be measures of many of the same factors that were identified in the present analysis and which were originally expected to assist in the determination of those factors, actually contributed very little to their determination. Only the Verbal Concepts test was useful in this respect, while the others tended to cluster with the general ability tests and to have loadings on the General Test factor. The factorial impurity of the Spatial Relations, Logical Reasoning, and Numerical Reasoning subtests of the California may be due to the fact that each of these subtests actually requires the testee to perform two somewhat different kinds of tasks, both of which are supposed to be satisfactory and highly correlated measures of the ability in question. That the combination of the scores for these separate tasks into a so-called "factor" score seems to result in an index that is more closely related to general mental ability than to any of the primary mental abilities suggests that the two separate tasks which are involved in the "factor" subtests are either not highly correlated or are too general to be valid measures of a primary ability.

Even though the factors that were extracted from the correlation matrices for the two status groups were essentially the same, the factorial structure which was obtained after the rotation of reference axes was much more clear-cut and definitive for the

High Status Group than for the Low Status Group. This is indicated by the fact that the tests that determined a given factor generally had higher loadings on that factor for the High Status Group and yet had projections on other axes that were distributed much more closely around the zero point of these axes. The oblique solution for the High Status Group seemed to meet the requirements of "simple structure" more adequately than that for the Low Status Group, and the final placement of the reference axes was somewhat more indeterminate for the Low Status Group. For the High Status Group this apparently betokens a state of affairs in which the tests tend to cluster closely together to determine specialized ability groupings that differentiate the total test space into rather clear-cut common factor spaces. For the Low Status Group this differentiation does not exist in nearly as definitive a fashion. This is observed most clearly when the various factors are plotted against one another by graphical methods. The lack of differentiation for the Low Status Group is then evident for all factors, but is particularly evident in the plots of V against R, WF against R, WF against V, V against GT, and GT against R.

The Verbal Meaning factor for the Low Status Group is a good example of a factor that is very poorly defined in comparison with its High Status counterpart. The tests that determine this factor for both status groups are the Vocabulary and Completion tests of the Chicago series and the Verbal Concepts test of the California, but the loadings for these tests are 0.62, 0.55, and 0.57, respectively, for the High Status Group and 0.40, 0.23, and 0.21, respectively, for the Low Status Group. This is rather unusual in view of the important rôle that verbal ability is supposed to take in determining mental test performance for the Low Status Group. A reasonable explanation for this rather appreciable difference in factor loadings is that verbal skill is an extremely generalized function for the Low Status Group and thus permeates all mental tasks to such a degree that it loses its factorial identity in the first order and becomes almost collinear with a general ability factor of the second order.

Evidence in favor of this interpretation is found in Tables V and VI, which show the correlations between group centroid axes which resulted when multiple group analyses were undertaken on the correlation matrices for both status groups. The same tables also show the projection of each axis on the second-order general

TABLE V—CORRELATONS BETWEEN GROUP CENTROID AXES AND LOADINGS ON SECOND-ORDER GENERAL FACTOR: LOW STATUS GROUP

Factor	Factor					
	V	WF	S	R	N	GT
V (Verbal Meaning)						
WF (Word Fluency)	79					
S (Space)	58	45				
R (Reasoning)	77	61	60			
N (Number)	72	68	51	68		
GT (General Test)	88	72	67	71	77	
r_{mg} (2nd-Order Loadings)	93	78	65	82	81	94

factor. The tests that were chosen for the various groups followed rather closely the factorial structure revealed in the centroid analyses, but an effort was made to include at least three tests in each group. The Number factor was represented by the Addition and Multiplication tests of the Chicago series and the Numerical Reasoning test of the California. The Verbal Meaning factor was represented by the Vocabulary and Completion tests of the Chicago series and the Verbal Concepts test of the California. The Space factor was represented by the Figures and Cards tests of the Chicago series and the Spatial Relations test of the California. The Word Fluency factor was represented by the First Letters, Four-Letter Words, and Suffixes tests of the Chicago series. The

TABLE VI—CORRELATIONS BETWEEN GROUP CENTROID AXES AND LOADINGS ON SECOND-ORDER GENERAL FACTOR: HIGH STATUS GROUP

Factor	Factor					
	V	WF	S	R	N	GT
V (Verbal Meaning)						
WF (Word Fluency)	58					
S (Space)	41	37				
R (Reasoning)	67	50	51			
N (Number)	59	60	52	60		
GT (General Test)	85	45	58	76	67	
r_{mg} (2nd-Order Loadings)	83	63	60	81	79	90

Reasoning factor was represented by the Letter Series, Letter Grouping, and Pedigrees tests of the Chicago series. And the General Test factor was represented by the Logical Reasoning test of the California and the Otis Beta and Henmon-Nelson general intelligence tests. These same groups were employed in the multiple group analyses for both status groups. The factor matrices which were obtained through this procedure did not yield any additional data beyond that already obtained from the centroid analyses, but the correlations between the group centroid axes which were secured as a by-product of this method revealed some very interesting relationships.

In Tables V and VI it can be observed that all the correlations except two are higher for the Low Status Group, and these discrepancies are particularly large for the Verbal Meaning and Word Fluency factors. In fact, the correlations of the Verbal Meaning factor with the other factors are so high for the Low Status Group that this factor has a loading of 0.93 on the second-order general intellective factor, making the one almost collinear with the other. This loading is about as high as that for the General Test factor, establishing rather convincingly the importance of verbal facility in the determination of tested intelligence for the Low Status Group. All of the other factors at the Low Status level also have higher loadings on the second-order general factor than their High Status counterparts. Apparently each of the primary mental abilities is more a function of a second-order parameter for the Low Status Group than for the High Status Group, and this second-order parameter is more thoroughly saturated with verbal components for that group. This suggests that a certain minimum level of verbal competence must be attained before a test of any special ability can become a more specific measure of that primary function, and for many Low Status youngsters this level has not been attained. All tests, both general intelligence tests and primary ability tests, are then more heavily weighted with the verbal component for the Low Status Group and would to this extent be less valid as measures of the particular ability or combination of abilities they purport to measure. This would be true, of course, of all tests except those specifically designed as tests of verbal competency.

This tendency for the primary mental functions of the Low Status Group to cluster around a general-verbal axis of the second

order is again indicative of the relative lack of differentiation that exists in the organization of mental abilities for the Low Status Group. This lack of differentiation has been manifest in three ways: in terms of the lack of definitiveness of the factors extracted, in terms of the amount of variance accounted for by the first centroid, and now in terms of the relative lack of independence among the primary abilities and the concomitant tendency for these abilities to cluster around a general-verbal factor of the second order.

Other interesting comparisons can be made from the data presented in Tables V and VI. Particularly interesting are the relationships between the Word Fluency factor and the other factors for the two groups. Like the Verbal Meaning factor, the Word Fluency factor has a much higher loading on the second-order general intellective factor for the Low Status Group, showing an increase from 0.63 to 0.78. This increase in general factor loading is similar to conditions in the first order, where the three tests that determine the Word Fluency factor—First Letters, Four-Letter Words, and Suffixes—can be observed to have considerably higher loadings on the first centroid for the Low Status Group (Tables I and II). This same Word Fluency factor is also much more highly correlated with the General Test and Verbal factors for the Low Status Group, increasing from 0.45 to 0.72 for the former and from 0.58 to 0.79 for the latter. Word fluency would appear to be a relatively independent ability at the High Status level, but at the Low Status level it is much less differentiated from the verbal-general ability complex as defined by the first and second order general factors of this study. This lack of differentiation is accompanied by an interesting change in factorial character. For the Low Status Group the Word Fluency factor is most highly correlated with the Verbal Meaning factor and then with the General Test and Number factors. For the High Status Group its highest correlation is with the Number factor. Again it would seem that a minimum level of verbal competency is required before we can presume to measure the special abilities, for a child cannot be fluent until he has a considerable repertoire of words with which to be fluent, and after that his fluency is limited only by his speed of association, which is a requisite for success in number computation and word fluency alike.

The comparisons which have been made between the two status

groups were those which seemed most significant to the author, but there are other differences which might also be considered. Among these are differences in individual test loadings for the two groups. It might be asked why the Suffixes test helped to determine the Word Fluency factor for the Low Status Group but not for the High Status Group, or why the Logical Reasoning test of the California had a much higher loading on the first centroid for the Low Status Group, or why the Pedigrees test of the Chicago series had a much higher loading on the first centroid for the High Status Group. It is beyond the scope of the present paper to consider these minor differences in detail, but the reader is invited to interpret these differences in the light of any explanations that appear to be logical and justified in terms of the data presented.

It might be contended that the relatively undifferentiated organization of mental abilities found for the Low Status Group in this study was actually an artifact of either or both of two irrelevant conditions: (a) a greater range of ability for the Low Status Group, resulting in generally higher variances and correspondingly higher correlations for that group; or (b) the influence of general performance level on the structure of abilities, which, because it would be lower for the Low Status Group, would leave in question whether this lack of differentiation was associated with Low Status *per se* or merely with the lower level of general performance associated with the Low Status Group. Evidence against the former interpretation emerged from a comparison made between the standard deviations computed for both groups for all tests, in which three standard deviations were observed to be significantly higher for the Low Status Group compared to two that were significantly higher for the High Status Group. The latter two were both for tests in the Chicago series, which were the most instrumental in establishing the evidence for a lack of differentiation among mental abilities for the Low Status Group. Evidence against the latter interpretation arose as a result of comparisons that were made between the factor matrices of a Low and a High Status Group which were "matched" for Otis Beta IQ but were still comparable to the original status groups in terms of their average indices of social status. Although this procedure is not without disadvantages, the evidence from this comparison again seemed to indicate a greater differentiation of mental abilities for the

High Status Group and a corresponding lack of differentiation for the Low Status Group.

This differentiation of traits, which has been reported as lacking for the mental abilities of the Low Status Group, is certainly not a new concept in the field of developmental psychology. As a theoretical construct it has served to organize and give meaning to data in the areas of mental, physical, and emotional growth as well as to the growth process as a whole (*3*, *4*, *7*, *9*). It has made one of its most recent appearances in the form of a hypothesis purporting to be a developmental principle of mental growth. Proposed by H. E. Garrett, this hypothesis states that: "Abstract or symbol intelligence changes in its organization as age increases from a fairly unified and general ability to a loosely organized group of abilities or factors" (*6*).

Although there is somewhat conflicting evidence about the differentiation of abilities with age, most studies in this area seem to report trends in the direction stated by the hypothesis. Assuming for the moment that this hypothesis has been adequately confirmed, we can combine this principle with the results of the present study and conjecture that the differentiation of mental abilities is not only a function of increasing age but also a function of the quality of the environment and the type of attendant mental stimulation which constantly acts upon the individual during the growth process. Optimum mental stimulation would facilitate the maximum differentiation of mental abilities. Since Low Status youngsters are far less subject to the type of intellectual stimulation that is conducive to the development of the abstract, academic skills that are so highly valued by the middle class, they would therefore possess a relatively undifferentiated organization of mental abilities. Additional data in the form of comparative studies of the structure of abilities for other groups would help to clarify the nature of the specific characteristics within the environment that are the most important agents in the facilitation or inhibition of this process of differentiation.

SUMMARY

The purpose of this study was to compare the factorial organization of mental abilities for a well-defined High and Low Social Status Group. To accomplish this purpose a comprehensive bat-

tery of mental tests was administered to a large group of eleven- and twelve-year-old subjects. This battery consisted of the Addition, Multiplication, Vocabulary, Completion, Figures, Cards, First Letters, Four-Letter Words, Suffixes, Letter Series, Letter Grouping, and Pedigrees tests of the *Chicago Tests of Primary Mental Abilities* for ages eleven to seventeen; the Spatial Relationships, Logical Reasoning, Numerical Reasoning, and Verbal Concepts subtests of the *California Short-Form Test of Mental Maturity*; the *Otis Beta Quick-Scoring Mental Ability Test*; and the elementary form of the *Henmon-Nelson Test of Mental Ability*. The subjects were classified into the High or Low Status Groups according to their scores on the "Index of Status Characteristics," and the final analysis was made with one hundred and seventy-two subjects in the Low Status Group and one hundred and seventy-nine subjects in the High Status Group (*15*). Scores on the eighteen tests in the battery were intercorrelated for each group, and the correlation matrices so obtained were factored by means of Thurstone's complete centroid method (*14*). The resulting factor patterns for the two groups were then compared. A Multiple Group analysis was also undertaken in order to compare the correlations between the factors extracted.

Essentially the same factors were identified for both status groups: Number, Verbal Meaning, Space, Word Fluency, Reasoning, and a General Test factor for which all the general ability tests had high loadings. But although the factors were essentially the same, the organization of mental abilities was found to be much less differentiated for the Low Status Group. This lack of differentiation in the organization of mental abilities for the Low Status Group was evident in three principal ways:

(1) The first centroid factor, similar to Spearman's "g" and expressing the communality of function between the several variables employed, accounted for a greater proportion of common factor variance for the Low Status Group than for the High Status Group.

(2) The factors extracted for the Low Status Group were not nearly as clear-cut and definitive as those for the High Status Group, and there was a sharpness and clarity in the final "simple structure" solution for the High Status Group that was entirely lacking for the Low Status Group.

(3) There was a relative lack of statistical independence among the factors extracted for the Low Status Group (i.e., the factors were more highly intercorrelated), and there were correspondingly higher loadings for these factors on the second-order general factor.

The rôles played by the Verbal Meaning and Word Fluency factors in the factor patterns for the two status groups suggested that the lack of differentiation associated with the Low Status Group was at least partly the result of the increased saturation of the general factor with verbal components. This condition apparently arises as a result of the fact that all mental tests, including the tests of primary or special abilities, become more verbally weighted for the Low Status youngster because he has not attained the requisite level of verbal competence to execute easily the modest verbal tasks that are inevitably involved in even the most "nonverbal" of tests.

REFERENCES

1) A. Anastasi and J. P. Foley, *Differential Psychology*, pp. 512-520. New York, Macmillan, 1949.

2) B. Balinsky, "An Analysis of the Mental Factors of Various Age Groups from Nine to Sixty," *Genetic Psychology Monographs*, 23: 191-234, 1941.

3) K. M. B. Bridges, "Emotional Development in Early Infancy," *Child Development* 3: 324-334, 1932.

4) G. E. Coghill, "Individuation Versus Integration in the Development of Behavior," *Journal of General Psychology*, 3: 431-435, 1930.

5) H. E. Garrett, A. I. Bryan, and R. E. Perl, "The Age Factor in Mental Organization," *Archives of Psychology*, No. 176, 1935.

6) H. E. Garrett, "A Developmental Theory of Intelligence," *American Psychologist*, 1: 372-378, 1946.

7) A. Gesell and H. Thompson, *Infant Behavior: Its Genesis and Growth*. New York, McGraw-Hill, 1934.

8) K. J. Holzinger and F. Swineford, "A Study in Factor Analysis: the Stability of a Bi-factor Solution," *Supplementary Educational Monographs*, No. 48. Chicago, University of Chicago Press, 1939.

9) M. Sherman, "The Differentiation of Emotional Responses in Infants," *Journal of Comparative Psychology*, VII: 265-284, 335-351, 1927.

10) F. Swineford, "Growth in the General and Verbal Bi-factors from Grade VII to IX, *Journal of Educational Psychology*, 38: 257-272, 1947.

11) F. Swineford, "General, Verbal and Spatial Bi-factors after Three Years," *Journal of Educational Psychology*, 40: 353-360, 1949.

12) W. H. Taylor and J. G. Darley, "Intercorrelations of the Test Scores Made by Selected Samples of Unemployed Men and Women," *Univ. Minn., Empl. Stab. Res. Inst. Bull.*, 3: 203-207, 1934.

13) G. Thomson, *The Factorial Analysis of Human Ability*, pp. 122–124. London, University of London Press, 1951.

14) L. L. Thurstone, *Multiple Factor Analysis*, pp. 149–176. Chicago, University of Chicago Press, 1947.

15) W. L. Warner, M. Meeker, and K. Eells, *Social Class in America*. Chicago, Science Research Associates, 1949.

EFFECT OF IMMEDIACY OF KNOWLEDGE OF CORRECTNESS OF RESPONSE UPON LEARNING

ARTHUR C. F. GILBERT

University of Nebraska

One of the existing problems in education is that of making the testing period a profitable learning experience for students. Some instructors have followed the practice of discussing the test items with the students as soon as possible after the completion of an examination. Others have provided the students with "model" responses similar to those involved in an examination. Still other instructors have given their attention to developing devices which permit the student to determine the correctness of his response as soon as the responses have been made.

The devices provided by the latter group of experimenters have been many and ingenious. Among such devices is the Science Research Associates (S.R.A.) Self-Scorer invented by Angell and Troyer (*2*). This device was made available commercially in 1949 and is designed for use with objective tests of either the multiple choice, matching or true-false variety.

The S.R.A. Self-Scorer consists of four parts, a combination answer sheet and key, a perforated mat, and a plastic frame or envelope. In the assembly of the Self-Scorer, the perforated mat is inserted between the answer-sheet and key, and this assembly is then inserted in the plastic envelope. In using the Self-Scorer the student punches a spot on his answer sheet corresponding to what he believes to be the correct response. If his selection of responses is correct a red dot appears on his answer sheet, but if his selection of response is incorrect, the answer sheet remains white at that place.

A review of the literature indicates that research using self-scoring devices has been going on for at least thirty years. Pressey (*7*) in 1926 reported the development of a self-scoring device which kept the problem before the student until he responded correctly. Among the research aimed at increased efficiency in learning through immediacy of knowledge of correctness of response are studies by Little (*5*), Peterson (*6*) and Stephens (*8*). Studies

involving the use of the Self-Scorer were made by Angell (*1*), Jones and Sawyer (*4*), Fields (*9*) and Blood (*3*). A consideration of previous studies in this area has indicated the need for a study which would test certain hypotheses within the framework of one research design while using the same group of subjects.

The purpose of this study, therefore, was to investigate the following three implications: (1) the effect of the S.R.A. Self-Scorer compared with the traditional method of testing followed by class discussion of the test items, when gain from two administrations of the same test is used as a criterion and with scholastic aptitude and achievement scores obtained from a preregistration test battery held constant; (2) the effectiveness of the S.R.A. Self-Scorer evaluated under the same conditions but with the score on a separate test as the criterion; (3) the differential effectiveness of the two methods of testing as regards learning outcomes when the test items are classified according to the educational outcomes which they are designed to measure.

PROCEDURE

Students of the beginning course in educational psychology at the University of Nebraska were used as subjects for the investigation. Subjects from eight sections, four instructors teaching two sections each, were used. One of the two sections taught by each instructor was designated as belonging to the experimental group, and the other section was used for the control group. In each of the experimental and the control groups there were four sections. There were ninety-eight students in this experimental group and one hundred and fourteen in the control group.

At the end of the first six weeks of the semester an examination covering the instruction up to that point was administered. This test was composed of seventy-four objective type items of the multiple choice and matching variety. The experimental group used the Self-Scorer, whereas the control group used standard type answer sheets.

The six-weeks examination was administered to the experimental and the control group in two different forms. Form A covered the entire six weeks of instruction and was administered to the control group during a fifty-minute class period. Form B contained the identical items of Form A but was divided into two parts; Part I

contained the first thirty-six items of Form A, and Part II contained the subsequent thirty-eight items.

The control group was allowed one fifty-minute period for the completion of the total test. The next class period was utilized for discussion of the test items. At this time the test booklet (form A) was returned to the students, together with their answer sheets. The instructors advised the students that notes were not to be taken during this period. After discussion of the test items all test forms and answer sheets were returned to the instructors.

The first part of the six-weeks test (Form B) was administered to the experimental classes during the first class period; the second class period was used for the remainder of the test by the experimental group; there was no discussion of any kind with this group.

Prior to the administration of the test to the experimental group the nature of the Self-Scorer was explained. Students were advised that the device was so constructed as to permit them to know whether they were responding correctly to the items while taking the test. The procedure to be followed while taking the test was explained.

Item analyses were made of all items in this first test. The total number of persons responding incorrectly to each item was tabulated and a post-test was constructed involving those concepts measured by the items most frequently missed in the original test. Four weeks later the original test was again administered to both the experimental and the control groups, together with the twenty-five-item post-test. Standard answer sheets were used by both groups for the second testing period. Both groups were tested without any warning on the same day.

The data from this investigation were analyzed by the analysis of covariance. Two criteria were employed, the first criterion being the gain on the six-weeks examination and the second criterion the score on the post-test. The control variables used were the scores on a locally constructed preregistration English Placement examination, the scores on the Linguistic subtest and the scores on the Quantitative subtest of the American Council on Education (ACE) Psychological Examination.

Items of the test were classified according to type of behavior measured in terms of course objectives and responses to each item tabulated for both administrations of the same test for both the

experimental and the control group. These data were used to determine the effect of the Self-Scorer on learning when different types of learning outcomes are considered.

RESULTS

Gain as a criterion. The first criterion used was the gain in scores on the second administration of the same test. The analysis of covariance was used to test the significance of the difference between the means of the two groups holding the control variables constant. In Table I are shown the results of the analysis of covariance. The value of F of 13.93 with 1 and 207 degrees of freedom is significant at the one per cent level of confidence.

By comparison of the two adjusted criteria means it may be seen that the students in the control group are favored over those in the experimental group. Therefore, it may be concluded that, insofar as no other pertinent factors contribute a bias, the evidence indicates that the gain achieved by the students in the control group

TABLE I—SIGNIFICANCE OF THE INFLUENCE OF THE S.R.A. SELF-SCORER ON THE GAIN ON THE SAME TEST WITH CONTROLS (ANALYSIS OF COVARIANCE)

Source of Variation	Degrees of Freedom	Residuals	
		Sum of squares	Mean square
Total	208	7115.50	
Within subgroups	207	6666.88	32.2071
Difference	1	448.62	448.62

$$F_{1,207} = \frac{448.62}{32.2071} = 13.93$$

$$t = \sqrt{F} = \sqrt{13.93} = 3.73$$

Adjustments:

$$\bar{Y}_E = (\bar{X}_{1E} - \bar{X}_{2T})a_1 + (\bar{X}_{2E} - \bar{X}_{2T})a_2 + (\bar{X}_{3E} - \bar{X}_{3T})a_3$$
$$= .37a_1 + .33a_2 + .12a_3$$
$$= .213$$

$$\bar{Y}_c = (\bar{X}_{1c} - \bar{X}_{1T})a_1 + (\bar{X}_{2c} - \bar{X}_{2T})a_2 + (\bar{X}_{3c} - \bar{X}_{3T})a_3$$
$$= -.32a_1 + (-.27)a_2 + (-.10)a_3$$
$$= -.178$$

$a_1 = .26152626$ $\bar{Y}_E = 1.79 - .21 = 1.58$

$a_2 = .29826304$ $\bar{Y}_C = 4.29 + .18 = 4.47$

$a_3 = .14161260$

was significantly greater than the gain achieved by those students in the experimental group. Thus, the hypothesis "there is no significant difference between the two group means for the second administration of the same test with the English Placement scores, the L scores and the Q scores controlled," is rejected.

The significance of the relationship between the criterion and the control variables in this investigation was tested by the analysis of regression. The obtained F value of 1.11 with 3 and 208 degrees of freedom is not significant. It may also be concluded that the obtained multiple correlation coefficient $Ry(x_1x_2x_3)$ of 0.126 is not significantly different from zero. Thus, it may be concluded that evidence is provided that the control variables did not account for a significant amount of variation in the criterion beyond that accounted for by the experimental-control group classification. To the extent to which validity is characteristic of the measurements employed, evidence from the analysis of results with gain as a criterion indicates that the students using the traditional answer sheets in the testing situation followed by class discussion at a later date achieve significantly higher when the test is readministered than students using the Self-Scorer without the benefit of discussion.

Post-test as a criterion. The second criterion used in determining the influence of the Self-Scorer on learning was the score on a twenty-five-item test which will be referred to hereafter as the post-test. As in the analysis of the results with gain as a criterion the analysis of covariance technique was employed for those data.

TABLE II—SIGNIFICANCE OF THE INFLUENCE OF THE S.R.A. SELF-SCORER ON THE SCORES ON THE POST TEST WITH CONTROL (ANALYSIS OF COVARIANCE)

Source of Variation	Degrees of Freedom	Residuals	
		Sum of squares	Mean square
Total	208	1327.86	
Within subgroups	207	1309.62	6.33
Difference	1	18.24	18.24

$$F_{1,207} = \frac{18.24}{6.33} = 2.88$$

$$t = \sqrt{F} = \sqrt{2.88} = 1.70$$

In Table II are shown the results of the second analysis of covariance. The value of F of 2.88 with 1 and 207 degrees of freedom is not significant. Therefore, the hypothesis, "there is no significant difference between the two group means for the scores on the post-test with the English Placement scores, the L scores and the Q scores controlled," is not rejected. It has not been possible to demonstrate that other than a single statistical population is represented by both these groups.

The significance of the relationship between the post-test as a criterion and the control variables was tested by the analysis of regression. An F value of 6.90 with 3 and 208 degrees of freedom was obtained. This is a greater value than that required at the one per cent level. Therefore, it may be concluded that the multiple correlation coefficient $Ry_1(x_1x_2x_3)$ of 0.50 is significantly different from zero. Evidence of the effectiveness of the controls in eliminating individual differences in the characteristics used as controls in this analysis is provided by the rejection of the zero correlation hypothesis.

To the extent to which validity is characteristic of the measurements employed, evidence from the analyses of results with the post-test scores as the criterion indicates that the contribution of the use of the Self-Scorer is not significantly greater than the traditional approach for students similar to those participating in the experiment under the same conditions.

Influence on different types of learning. Test items in the seventy-four-item test were classified into three categories: (1) definition items which involved the definition of psychological terms presented in class, (2) items involving the knowledge of facts, and (3) items requiring application of principles. A comparison of the responses of both the experimental and the control groups on both administrations of the same test was made for the three types of items. The responses were tabulated to determine the percentage of subjects in each group who responded incorrectly to the item on the first administration of the test but correctly the second time.

Analysis of the responses to the different types of test items shows that the experimental group had an equal or higher percentage of successes on four out of ten, or 40.0 per cent of the items which involved the definition of terms. Of the items which involved the knowledge of facts the experimental group had an equal or higher percentage on ten out of twenty-six, or 58.5 per cent of the items.

On the items which involved application of principles the experimental group was equal or higher percentage-wise on fifteen out of thirty-eight, or 39.5 per cent of the items.

No significant difference was found between the paired percentages for each of the three types of items. From the results of this analysis evidence is provided that the S.R.A. Self-Scorer does not significantly influence learning on one type of item more than others according to the classification of items used in this study, definitions, knowledge of facts and application of principles.

DISCUSSION

In this study it was discovered that, in general, students actually required twice the length of time to complete a test by use of the S.R.A. Self-Scorer than by other methods. Therefore, from the viewpoint of administration the use of the S.R.A. Self-Scorer would have to be justified by the fact that those using this device could perform equally with those tested in the traditional method in which one period is used for testing and the next one being used for discussion.

The results of this study indicated that when gain on the same test was used as a criterion those using the standard answer sheets accompanied by discussion performed significantly better than those using the S.R.A. Self-Scorer. If the experimental group had more experience with the S.R.A. Self-Scorer, it is possible that they might have shown a higher performance. On the twenty-five-item test there was no significant difference between the two groups. If subjects similar to those in this experiment consistently perform in this manner when similar methods of testing are used, then the use of the S.R.A. Self-Scorer method would not be any more justified than the traditional method.

SUMMARY

The purpose of this study was to determine the effect of the S.R.A. Self-Scorer upon learning when used in a testing situation as compared with the use of standard answer sheets when the latter method of testing is accompanied by class discussion of the test items.

Two criteria were used in the measurement of the outcomes of this experiment. Both criteria consisted of objective type examinations administered four weeks after the original learning experience.

The first criterion was the gain achieved on the second administration of a seventy-four-item test and by use of this criterion in the analysis, the results of the investigation indicated that the control group performed significantly better than the experimental group at the one per cent level of confidence. The analysis of covariance used in the analysis yielded an F value of 13.93 with 1 and 207 degrees of freedom.

An analysis of regression was performed to test the significance of the relationship between the first criterion and the control variables. The F value of 1.11 with 3 and 208 degrees of freedom provided evidence that the control variables were not effective in explaining a significant portion of the variation in the gains criterion.

The second criterion was the scores on a twenty-five-item test constructed around concepts measured by the twenty-five most difficult items of the seventy-four-item test. With this post-test as a criterion, analysis of the results showed no significant difference between the experimental and the control groups. Analysis of regression yielded a value of F of 6.90 with 3 and 208 degrees of freedom, which provides evidence that the control variables were effective in eliminating a significant portion of the criterion variation attributable to individual differences used as controls in the experiment.

An analysis of items of the seventy-four-item test was performed comparing responses on both administrations of the test. The results of this analysis indicated that the S.R.A. Self-Scorer does not significantly affect learning on items measuring one type of learning outcome (e.g. definition, knowledge of facts or application of principles) more than on items measuring other types of learning outcomes.

REFERENCES

1) George W. Angell, "The Effect of Immediate Knowledge of Quiz Results on Final Examination Scores in Freshman Chemistry," *Journal of Educational Research*, 42: 391–394, 1949.

2) George W. Angell and Maurice E. Troyer, "A New Self-scoring Device for Improving Instruction," *School and Society*, 67: 84–85, 1948.

3) Donald F. Blood, "The Effect of the S.R.A. Self-Scorer in the Measurement of Spelling Ability," Unpublished doctoral dissertation, State University of Iowa, 1951.

4) H. L. Jones and M. O. Sawyer, "A New Evaluation Instrument," *Journal of Educational Research*, 42: 381–385, 1949.

5) James K. Little, "Results of the Use of Machines for Testing and Drill upon Learning in Educational Psychology," *Journal of Experimental Education*, 3: 45–49, 1934.

6) J. C. Peterson, "A New Device for Teaching, Testing and Research in Learning," *Transactions of the Kansas Academy of Science*, 33: 41–47, 1930.

7) Sidney L. Pressey, "A Simple Device Which Gives Tests and Scores—and Teaches," *School and Society*, 23: 575–576, 1926.

8) J. M. Stephens, "The Effect of Symbolic Punishment upon Strong and upon Weak Associations," *The Journal of General Psychology*, 25: 177–185, 1941.

9) Maurice E. Troyer and George W. Angell, *Manual for the S.R.A. Self-Scorer*. Chicago, Science Research Associates, 1949.

EFFECT OF TEACHER-STUDENT INTERVIEWS ON CLASSROOM ACHIEVEMENT[1]

ARTHUR J. HOEHN and ELI SALTZ

Air Force Personnel and Training Research Center,

Chanute Air Force Base, Illinois

Many psychologists and educators have encouraged teachers to take on counseling-interview functions (*1–3, 6, 11*). Few studies are reported in the literature, however, to indicate the effectiveness of such interviews. At present there is very little evidence concerning what kinds of interviews are effective for improving classroom performance, or what types of students benefit most from interviews. In addition, the studies reported in the literature usually employ the experimenter himself as interviewer. It is probably a safe assumption that most experimenters in this area are particularly interested in the effect of personal, teacher-student relations, and are consequently good interviewers.

The present series of studies was designed to investigate the following questions:

(a) Can the average teacher in an Air Force technical training course improve student performance by means of interview techniques?

(b) Can Air Force students be identified who will benefit most from an interview with their teacher?

(c) Can particularly effective interview techniques be identified?

Two of the better controlled experiments reported in the literature are those of Sherriffs (*8*) and Bradt and Duncan (*3*). Both studies report success in raising student grades by means of interview techniques. Both of these studies involved only the experimenters as interviewer. Sherriffs found that interviews were most effective with students rated as being high in "self tension," "family tension," and "social tension" on the basis of the interview material.

[1] This research was carried out under the Air Force Personnel and Training Research Center, Chanute Air Force Base, Illinois, in support of Project 7714, Task 77249. The writers wish to express their appreciation to Dr. Henry Adams, Mr. Charles E. Zeleny, and M. Sgt. Robert C. Evans for their assistance in designing and implementing the present research.

Bradt and Duncan utilized a thirty-minute interview, Sherriffs a sixty-minute interview. Bradt and Duncan characterized their most effective interview procedure as one in which the interview was oriented about a discussion of the student's goals and ambitions, in contrast with discussion of world affairs or no interview. Sherriffs characterized his most effective interview procedure as one in which the interview was oriented around the student's life history, interests, and background.

In a somewhat related study, Briggs and Roe (*4*) report that encouraging airmen to "gripe" had a beneficial effect on the morale of the airmen.

Experiment I attempted to establish a teacher-student relationship similar to that described by Sherriffs, Bradt and Duncan, and Briggs and Roe. As in the Bradt and Duncan and the Sherriffs studies, teachers attempted to discuss the personal problems of the students. As in the Briggs and Roe study, students were encouraged to "gripe" about problems that bothered them. The writers hypothesized that anxiety and rigidity (as defined by tests described later in this paper) should reflect the types of tension described by Sherriffs. Therefore, anxious and rigid students should be helped by the catharsis opportunities of the interviews more than nonanxious and nonrigid students.

EXPERIMENT I

Method

Subjects. Forty-eight students were selected for interviewing out of a group of two hundred and seven students in their first four weeks of the Air Force Weather Observers School, the remaining one hundred and fifty-nine students comprised the control group. The two hundred and seven students comprised two classes, approximately equal in size, which entered the course within four weeks of each other. The course is a complex, technical one, and the majority of students in it are college caliber.

Each entering group was under a different set of two head teachers who did a large part of the teaching. These four men did all the interviewing in Experiment I.

Students chosen for interviewing were randomly selected from a population which had been stratified with regard to anxiety and rigidity. Anxiety was defined by the Taylor Manifest Anxiety Scale (*9*), a questionnaire device composed of MMPI items selected

by clinicians as indicators of manifest anxiety. Most of the items deal with symptoms of overt anxiety, e.g., sweating. Rigidity was defined by the Wesley Rigidity Scale (*10*), constructed in a manner similar to that used by Taylor. The items on this scale deal largely with persistence, constriction, and moralistic behavior patterns. Students were considered anxious or rigid if they obtained scores at or above the 80th centile on the respective tests. Students obtaining scores below the 80th centile were considered nonanxious or nonrigid.

Teacher training. The teachers were trained in the interview technique in two groups of two. Each group was trained for one hour on each of three days. On the first day of training, teachers were oriented as to the sorts of interviews they were to conduct. An opening paragraph for use with students was supplied to the teachers in which the teacher was to point out that, due to the size of the classes, it was difficult for teachers to get to know students. Teachers were then to ask questions about the student's marital status, how far from home the student was, and how the student liked the course. When a comment by the student suggested a problem area, this area was to be probed. Following this, the teachers were to follow the students in whatever topics the students raised. Teachers were told to encourage students to gripe about conditions on the base, at home, etc., and to take a permissive attitude toward such griping. During the remainder of the first training session, a psychologist demonstrated the technique using an airman assigned to AFPTRC as interviewee. The demonstration was then discussed with teachers to point up crucial facets of the interview. On the second and third days the teachers practiced interviewing AFPTRC airmen and these practice interviews were discussed.

Interviews lasted approximately thirty minutes and were conducted within the first two weeks of the course for most students. All were completed before the end of the third week.

Criteria. Grades for the first phase of the course were determined on the basis of an objective test given at the end of the fourth week of the course. Because of the high standards set for this course, approximately twenty-five per cent of the students in every entering class fail the first phase of the course and must repeat the phase. The decision on failures, in the two classes studied, was

made by an administrative officer who did not know which students had been interviewed, and was made on the basis of phase grades.

Results

An analysis of variance using phase grades indicated no differences between interviewed and noninterviewed students, and no greater effect of interviews for anxious or rigid as opposed to nonanxious or nonrigid, respectively.

Analysis of failure rates suggests that the interviews may have had some effect on students who were borderline between passing and failing. However, the results are not all in the predicted directions.

It had been anticipated that interviewed airmen would do better in classwork than airmen who were not interviewed. Failure data do not show any *general* tendency toward less failure on the part of interviewed as compared with not interviewed airmen. Interview effects seem to have been specific to personality types. For example, rigid students who were interviewed actually had a *higher* washback rate than those who were not interviewed. This difference, shown in Figure 1, is significant at only about the 0.10 level. As can also be seen in Figure 1, the data do not sustain the prediction that interviews would be more beneficial for rigid than for nonrigid

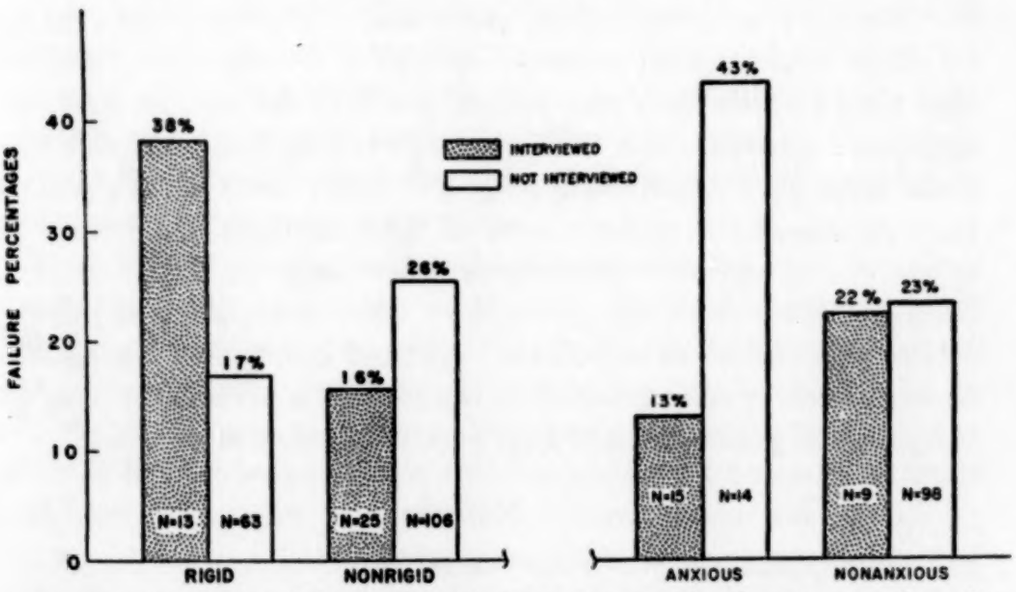

FIGURE 1. Effect of "gripe" interviews on washback rate as a function of rigidity and of anxiety.

students; the difference between the increase in failure rate for interviewed rigid students and the decrease in washback rate for the interviewed nonrigid students is significant at beyond the 0.05 level. These results suggest that rigid students should probably not be interviewed, at least not by the procedures described in this experiment.

The anticipation that interviews would be beneficial to anxious students tends to be sustained, as can be seen in Figure 1. The percentage of failures among interviewed anxious students is considerably less than the percentage of failures among not interviewed anxious students; the difference between percentages was significant at beyond the 0.06 level.

EXPERIMENT II

This experiment was planned as an extension of Experiment I. In the present experiment a larger number of teacher interviewers was employed and a second interview technique was developed with the hope it would prove effective for rigid students. Also, a more complex, factorial design was used in this experiment than in the first; this should increase the sensitivity of statistical analysis.

Clinical theory suggests a possible explanation for the opposite effects of "gripe" interviews on anxious students and rigid students. For anxious students the gripe interviews might have had a catharsis effect that reduced anxiety. However, clinical theory suggests that rigid persons have as a central problem the conflict over their aggressive impulses and that expression of aggression floods them with anxiety. Consequently, "gripe" interviews might actually have increased the anxiety level of rigid students. Therefore, two types of "interview" procedures were used in Experiment II. Some students received the "gripe" interviews of Experiment I, others received a "satisfactions" oriented interview. The satisfactions interviews concentrated on the present sources of satisfaction, the personal goals, and the hopes for the future of students.

Method

Experimental design. Four variables were manipulated in a 3 x 3 x 3 x 2 factorial design with three students in each of the fifty-four cells. The independent variables were:

(a) Anxiety: Students were stratified into three levels of anxiety

on the basis of scores on the Taylor Manifest Anxiety Scale (*9*). High anxious were students at or above the 80th centile. Medium anxious were students between the 21st and 79th centiles. Low anxious were students at or below the 20th centile. It should be noted that the low anxious group in Experiment I was the equivalent of the combined middle and low anxious groups in Experiment II.

(b) Rigidity: Students were stratified on the basis of the Wesley scale (*10*) into three levels of rigidity similar to the three levels of anxiety. As was true of anxiety, the low rigid group in Experiment I was the equivalent of the combined middle and low rigid groups in Experiment II.

(c) Interview technique: Students were placed in one of three interview groups—"gripe," satisfaction, or not interviewed.

(d) Aptitude: Students were stratified into high and low aptitude for the course on the basis of the Electronic Specialty Aptitude Index. Students with stanines of 6 and 7 were in the low group; those with stanines of 8 and 9 were in the high group.

Subjects. Students were one hundred and sixty-two airmen in the first phase of the Air Force Armament System Fundamental course, a course primarily concerned with fundamentals of electronics. These one hundred and sixty-two students were selected from approximately six hundred entering students. Selection was random within the stratifications determined by the individual difference variables being studied and was accomplished before the beginning of the first phase of the course. All students were above the 70th centile on the aptitude index used, and so the group aptitude is probably not lower than that found in the freshman class of most large state universities.

The interviews were conducted by thirty-one teachers. Teachers interviewed only their own students. Assignment was random with the restrictions that not more than five students requiring interviewing were to be assigned to any one teacher and that no teacher was to have more than one student from any of the fifty-four cells in the design. These restrictions were necessary in order to randomize teacher differences in interviewing and teaching ability. Teachers differed greatly in background. Approximately half were civilian and half were airmen.

Interviews. All interviews took place within the first three days of the first phase.

Each teacher attended these training sessions, conducted as follows.

Session 1: This was a lecture and discussion period approximately one hour in duration and covering the background, purpose and general nature of the study, the rôle which the teachers would play in the study, the kinds of interviews to be given, and the nature of the subsequent training. For this session, teachers met in groups of fifteen to twenty.

Session 2: This session, which teachers attended in groups of four to eight, included a detailed description of the interview procedures, a demonstration, and practice in interviewing. The teachers were told that, in interviewing, they should: (a) create the impression of being genuinely interested in the student, (b) induce the student to choose the topics dealt with and to speak freely in relation to them, (c) exert influence on the content of discussion only insofar as necessary to maintain the differentiation between "gripe" and "personal interest" interviews, (d) adopt an acceptant attitude but avoid expressions of personal approval or disapproval of the students' views, (e) give no advice and make no promises of action aimed at solving the student's personal problems, and (f) conclude the interview on a supportive note.

Following this description of the interviewer rôle, a demonstration was given of both the "gripe" and the "personal interest" interview with teachers taking the rôle of students being interviewed. Subsequent to this, each teacher interviewed another teacher and his performance was critiqued by other members of the group.

Session 3: The third session was a "refresher" session which teachers attended not more than three days prior to the time they did their first interviewing in the experiment. The details of the interviewing were again discussed and any further questions answered. It was again emphasized that in the "personal interest" interview, teachers were to encourage subjects to talk about things they liked, and that in the "gripe" interviews, teachers were to encourage subjects to talk about things that were bothering them.

Two check lists were constructed as devices for obtaining information which teachers were to use in conducting the interviews. On one of these, students were asked to check or list their five principal interests or sources of satisfaction. On the other, students were asked to check or list their five chief "gripes."

All the teacher training was accomplished by one airman research technician. This research technician, an airman with extensive interviewing experience, had been intensively trained by the experimenters in conduct of the specific types of interviews used in the present study.

Interviews lasted thirty to forty-five minutes per student.

Criteria. An objective test of forty-five items was administered to students at the end of the first two-week phase of the course. The questions involved computation of voltages, voltage drops, resistances, etc. The split half reliability of the test was found to be 0.90.

No failure data were collected since only between five and ten per cent of students in the first phase of this course failed it.

Results

As in Experiment I, there is no evidence that interviews significantly affect average grades. The only significant F in the analysis indicated that persons with high aptitude learn more than those with low aptitude. This difference is significant at beyond the 0.005 level, $F = 9.79$ with 1 and 108 degrees of freedom.

While failure rate was too low to permit analysis, it was possible to get some idea concerning the effect of interviews on performance by the borderline achievement students. The lowest score in each cell for low ability students should reflect the performance of borderline students. This analysis reduced the total sample size

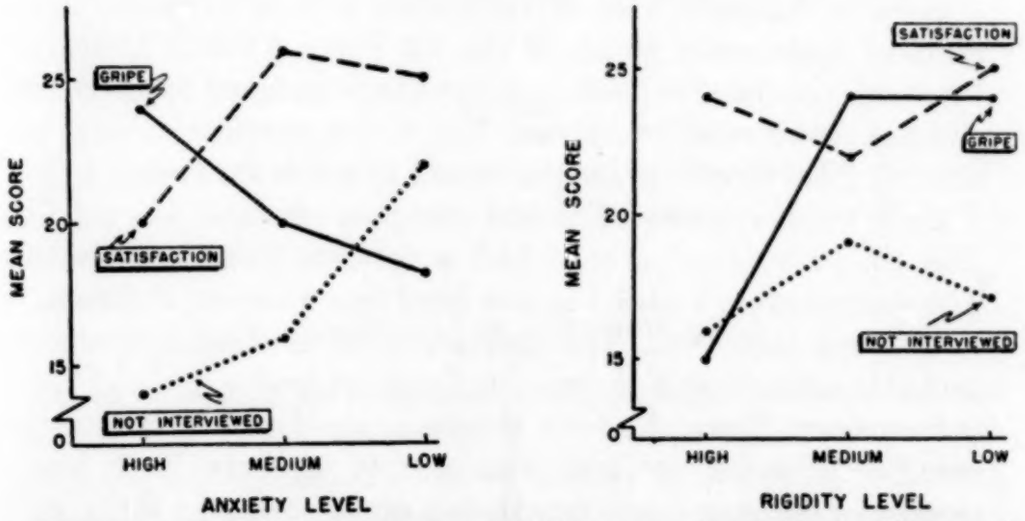

FIGURE 2. Effect of interview techniques on borderline achievement students at various levels of anxiety and of rigidity.

to twenty-seven and statistical analysis was not worthwhile. However, some of the trends may be of interest. Figure 2 indicates the effects of the various interview conditions on the borderline students. Again, as in Experiment I, anxious borderline students tend to be helped by gripe interviews. They are also aided, but to a lesser extent, by the satisfactions interviews. As in Experiment I, there is a tendency for low anxious students to be hindered by "gripe" interviews. There is, however, a tendency for low anxious students to be helped by satisfactions type interviews.

Figure 2 also indicates the effects of interviews on borderline achievement rigid students. As in Experiment I, gripe interviews tend to have a detrimental effect on rigid borderline students. Satisfaction type interviews appear to aid performance of the rigid borderline students.

In general, the two experiments are consistent in not indicating gains, due to interviews, by the average student.

EXPERIMENT III

This experiment was conducted at about the same time as Experiment II. It considers the effect of a group interview technique in which the students were encouraged to gripe. Each group was interviewed twice, as against the one interview in the previous experiment. Personality variables were not considered.

Method

Subjects. Subjects were seventy-seven airmen in phase 1, Mechanical Instrument phase, of the Air Force Aircraft Instrument Mechanics course. Students were randomly assigned to interviewed and not interviewed conditions. The six teacher-interviewers were selected from those regularly assigned to teach the course.

Each teacher interviewed one group of seven of his students. The six interviewed groups had a total of forty-two students. The control groups varied in size from four to seven students, the total being thirty-five. The aptitude level of these students was probably equivalent to a senior high school level.

Interviews. Three one-hour classes in nondirective group counseling techniques were held with the six teachers. Each teacher received a mimeographed description of the proposed study which also outlined his counseling duties. A "model gripe session" was

TABLE I—EFFECT OF GROUP INTERVIEWS ON GRADES

	Interviewed	Not Interviewed
Mean grade	47.30	46.40
S.D.	7.42	8.71
N	42.00	35.00

presented at the first class, followed by a short lecture on the group session procedure.

Both handout and lecture emphasized that the students should be encouraged to express complaints and resentments, especially those related to the course. The teachers were cautioned that they must be permissive and objective and must convince the students that no retaliation was possible.

The last two training sessions consisted of rôle playing. Each teacher was given an opportunity to act as discussion leader while the other teachers took the part of students. The sessions were tape recorded and these recordings were played back to the teachers and analyzed.

Each student group received one interview session during the first week and one during the second week of the three week phase. Each of these lasted approximately forty-five minutes.

Criteria. The measure of achievement in this study was the final phase examination. The Kuder-Richardson (Formula 20) reliability of this test was 0.76.

The failure rate in this course was too small to be of use as a criterion measure.

Results

The difference between phase grades for interviewed and non-interviewed students was small and not statistically significant, as can be seen in Table I.

SUMMARY AND CONCLUSIONS

Previous studies have reported that personal interviews between teachers and their students have resulted in significant average student grade increments. These results have been particularly striking when the interviews dealt with the personal interests and problems of the students. However, in most of the studies reported

the teacher-interviewer was a person whose background made him exceptionally suitable for conducting such interviews. The present series of three experiments tried to reproduce the facilitating results of teacher-student interviews using, as interviewers, a random sample of Air Force teachers. These teachers were given a three hour course in interview training that included lectures and rôle playing. The interview techniques taught were based on those that previous studies have reported as particularly effective.

The results suggest that the interviews influenced failure rates. However, this influence was not always beneficial. The results suggest that while *"gripe" oriented* interview techniques may *decrease* the failure rate for anxious students, these techniques may actually *increase* the failure rate for rigid students. On the other hand, interviews oriented toward *personal sources of satisfactions* appear to *decrease* the failure rate among borderline achievement rigid students. These results suggest that teacher interviews may be harmful in some cases.

All three experiments were consistent in finding that the *average grades* of interviewed students were not better than the average grades of students who were not interviewed. These results were consistent for *individual* interviews oriented toward student problems and "gripes;" for *group* interviews oriented toward student gripes; and for *individual* interviews oriented toward student goals, hopes for the future and sources of personal satisfaction. Nor did the student personality variables examined interact with interviews by the random sample of teachers. Interviews did not raise grade averages for anxious, rigid, nonanxious, or nonrigid students. It should be noted that Martin (*7*), too, found that interviews did not affect average grades but did affect number of course failures.

These results suggest that the average teacher, with only a minimal amount of training in interview techniques, cannot improve grade averages by means of interviews. Coupled with the possibility that such interviews may be harmful for some borderline achievers, these results indicate that caution should be exercised in recommending that teachers counsel their students.

Definite implications for future research emerge from these experiments: (a) The criterion of failure rate should be considered more carefully than it has in the past. More emphasis should be placed on the effects of interviews on borderline achievers. (b)

The sorts of teachers who can succeed in interviewing should be determined. Just as different interview techniques have different effects on students, depending on students' personality, teacher personality may influence the type of interview technique which a teacher can use successfully. It would be consistent with clinical intuition if future results showed that specific teachers were successful interviewers only when they used specific techniques, and used them only on students with specific personality characteristics. (c) The interaction between student personality and specific interview techniques should be investigated.

REFERENCES

1) D. S. Arbuckle, "Can English Teachers be Counselors?" *English Journal*, 42: 192–193, 1953.

2) D. S. Arbuckle, *Teacher Counseling*. Cambridge, Mass., Addison-Wesley Press, 1950.

3) K. W. Bradt and C. P. Duncan, "Degree of Personal Relationship between Instructor and Student as a Factor in Course Grade Improvement," *American Psychologist*, 6: 368 (abst), 1951.

4) L. J. Briggs and R. M. Roe, "Morale as a Function of Opportunity to Register Complaints," *Technical Report* HRRC-TR-53-4, Human Resources Research Center, Lackland Air Force Base, Texas, March, 1953.

5) C. A. Curran, "Guidance and Counseling in Education," *Education*, 73: 223–228, 1952.

6) C. S. Knape and E. A. Walsh, *Training Analysis and Development Information Bulletin*, 3: 8–12, 1952.

7) G. C. Martin, "Interviewing the Failing Student" *Journal of Educational Research*, 46: 53–60, 1952.

8) A. C. Sherriffs, "Modification of Academic Performance through Personal Interview," *Journal of Applied Psychology*, 33: 339–346, 1949.

9) Janet A. Taylor, "A Personality Scale of Manifest Anxiety," *Journal of Abnormal and Social Psychology*, 48: 285–290, 1953.

10) Elizabeth Wesley, "Perseverative Behavior in a Concept-Formation Task as a Function of Manifest Anxiety and Rigidity," *Journal of Abnormal and Social Psychology*, 48: 129–134, 1953.

11) G. C. Wrenn, *Training Analysis and Development Information Bulletin*, 4: 7–12, 1953.

A LONGITUDINAL STUDY OF SIBLING RESEMBLANCES IN INTELLIGENCE AND ACHIEVEMENT[1]

SARAH M. SCHOONOVER

The major problem of this study is to determine the amount of sibling resemblance in longitudinal growth records of mental ability and educational achievement. The following specific points have been investigated: (a) To what extent do children of the same family resemble each other in intelligence and in educational achievement? (b) To what extent are differences between siblings reduced as compared to those that exist between unrelated pairs within the same population? (c) To what degree do the sibling resemblances in intelligence and achievement correspond to each other?

Investigations concerning familial resemblances in mental ability have been numerous since the initial studies were made by Galton nearly ninety years ago. Studies of the resemblances of siblings in mental ability and in school achievement have been reviewed in the writer's doctoral dissertation (*2*). The present study has utilized the longitudinal approach to this problem in a manner that has not been employed with data analyzing sibling resemblances in intelligence or achievement.

SIBLINGS SELECTED FOR STUDY

Source of the data. The data were secured from the records of the University Elementary School at the University of Michigan. Of the children in this school, seventy per cent come from professional homes, while thirty per cent come from business and white-collar homes.

There are approximately seven applications for each vacancy at the University Elementary School. The primary purpose for which the school was established, i.e., the promotion of research

[1] This report is a portion of the writer's doctoral dissertation, "Sibling Resemblances in Achievement" (Supplemented by the mental-age data). See: *Dissertation Abstracts*, Vol. 13, No. 5, Publication No. 5726, Ann Arbor, University Microfilms, 1953. Acknowledgment is made to Dean Willard C. Olson, University of Michigan, who directed the original research. The author's present location is Ottawa, Ohio.

in child development, is the main guide in admission policies. Consideration is given to the suitability of children as subjects for growth studies and for the type of educational program provided. Applications are not accepted for the blind, the deaf, the seriously crippled, nor the severely mentally retarded. Preference is given to children likely to remain through sixth grade, to children seeking entrance at an early age, and to children having siblings already enrolled. An attempt is made to keep a balance between the number of boys and the number of girls at each age level.

The school endeavors to provide for the development of the child as a whole in his social setting. It takes into account physical hygiene, mental hygiene, pacing of growth, individual differences, and active participation in group life. Its curriculum is flexible and emergent, is based upon a concept of learning through experience, and is determined in the light of children's interests and present needs.

Description of the data. The psychology division of the University Elementary School maintains an intelligence test schedule on a yearly basis. The Stanford-Binet Revised Scale is administered annually within two weeks of each child's birthdate, or six months following the birthdate of children born in the summer. An achievement test schedule is maintained on a six-month basis. The Gates Reading Test is given semiannually in October and in April, until the child achieves a score of 102 months, at which time he is given the Stanford Achievement Test, Primary Battery. The Intermediate Battery is administered in grades four through six.

Records of intelligence and achievement test scores have been filed in the Psychometric Unit for all children who have attended and who have been tested at the University Elementary School. Since the school was opened in 1930, longitudinal material has been collected over a twenty-five-year period.

Selection of the data. All true sibling pairs, with chronological age overlap, and with four or more scores per sib on the Stanford-Binet Test and on the Stanford Achievement Test, from the fall of 1929 through the spring of 1951, were utilized in this study. With these qualifications, fifty-nine sibling pairs were found for intelligence; sixty-four pairs for arithmetic, education, reading, and spelling; forty-two pairs for literature and social studies; forty pairs for language; and thirty-eight pairs for science. There were fewer sibling combinations in language, literature, science, and social

studies because these tests are not found in the Primary Battery; therefore, there is less chance for a child to have four scores on these tests than for him to have that number on the other tests in the battery.

METHODS FOR THE COMPARISON OF LONGITUDINAL SIBLING RECORDS

The longitudinal approach. The longitudinal records, which involved measurement of the same children year after year, provided the data for this study. They have the following advantages: (a) Some of the variables found in the test situation, such as the child's disposition on any particular date or the child's reaction to being tested for the first time, are minimized. (b) Comparison of children at the same chronological age, rather than at dissimilar ages, is permitted.

For each family included in this study a mental growth graph and eight achievement growth graphs were constructed. This meant that three hundred and forty-four growth graphs and seven hundred and fifty-seven individual growth curves were plotted.

Chronological, mental, and achievement ages were expressed in months. Chronological ages were plotted on the abscissas of the graphs, while mental and/or achievement ages were plotted on the ordinates of the graphs.

Fitting the linear equation. The linear equation best fitting the data was found to eliminate the observed variation and to determine some constant rate of growth which may be used to characterize the observed results. The equation of a straight line, $y = ax + b$, which gives the slope and the intercept of the line used to describe the growth-age relationship, was found by the method of the least squares fit. The linear fit for each child for intelligence and for each of the eight achievement variables was plotted graphically.

The method of the mean of the average differences. For each pair of siblings the limits of the overlap of their chronological ages were found, and from these the midpoint of the overlap was computed. Using the linear best fit, the age scores of each pair of siblings were read at these midpoints. The difference between these two ages was found, yielding the average difference for each sibling pair. The sum of these average differences was divided by the number of pairs, to secure the mean of the sib average differences. In addition to the group of siblings as a whole, the above was calcu-

lated for brother-brother, sister-sister, and brother-sister combinations.

It may be noted that the difference found for each pair is the average difference for the whole overlap period for the pair, because of the nature of the straight line. Also, because the linear best fit was found by utilizing a number of test scores for each child, a longitudinal effect is reflected in the average differences.

The unrelated pairs. To secure comparable data for unrelated pairs, siblings of this study were paired randomly with a nonrelated partner, and the means of the average differences were calculated for the unrelated pairs. By utilizing the same children for the sibling and for the nonsibling pairs, all group variables (such as age, sex, IQ, EQ, physical factors, emotional factors, social maturity, home environment, and school environment) are neutralized, resulting in two groups, comparable to each other in all of these respects.

The ratios. Ratios were computed by using the means of the average differences of the sib-pairs as dividends and those of the corresponding unrelated pairs as divisors. The resulting ratios were subtracted from 1.00 and then were multiplied by one hundred, producing in percentages the extent to which the sibling differences were reduced beyond those of the nonrelated pairs. To determine the significance of the differences between the sib-pair means and the unrelated pair means a t test for significance was employed.

The method of correlation. In addition to the ratio of the means, correlation coefficients were computed for the sibling pairs. A Pearson product-moment formula was used, as was the method of intraclass correlation. To discover the significance of these correlations a t test of an observed correlation was employed.

SIBLING DIFFERENCE AND RESEMBLANCE

The mean of the average differences. The results obtained by the method of the mean of the average differences for sibling pairs and for unrelated pairs are given in Table I for intelligence and achievement.

The percentages of the reduction of the means of the average differences between the sibling and the nonrelated pairs in this study are given in Table II.

These siblings consistently were found to have smaller means of average differences than the unrelated children for the group

TABLE I—MEANS OF AVERAGE DIFFERENCES IN MENTAL AND EDUCATIONAL AGE SCORES, EXPRESSED IN MONTHS

Mental or Achievement Age Measure	Total Group	Boy-Boy	Girl-Girl	Boy-Girl
Mental				
Sibling	9.9	9.4	8.1	11.5
Unrelated	18.3	23.5	14.2	18.1
Arithmetic				
Sibling	10.7	12.2	9.2	10.6
Unrelated	15.1	17.7	10.1	13.8
Education				
Sibling	12.6	12.2	12.3	13.0
Unrelated	19.3	23.6	17.7	15.8
Language				
Sibling	20.7	16.1	23.3	22.2
Unrelated	30.1	31.8	36.5	28.5
Literature				
Sibling	14.3	10.0	16.7	15.6
Unrelated	20.7	18.2	22.7	17.4
Reading				
Sibling	16.4	13.0	17.8	18.3
Unrelated	25.9	32.6	22.4	26.1
Science				
Sibling	16.0	14.3	16.1	16.7
Unrelated	20.5	24.1	23.3	23.0
Social studies				
Sibling	13.6	12.3	9.7	17.6
Unrelated	19.7	25.2	13.8	21.8
Spelling				
Sibling	14.7	13.1	14.3	16.1
Unrelated	21.5	25.5	20.3	20.8

TABLE II—PERCENTAGES OF REDUCTION OF MEANS OF AVERAGE DIFFERENCES BETWEEN SIBLING AND NONSIBLING PAIRS

Mental or Achievement Age	Total Group	Boy-Boy	Girl-Girl	Boy-Girl
Mental	45.6	60.4	42.8	36.5
Arithmetic	29.2	31.1	8.9	23.2
Education	34.8	48.3	30.5	17.8
Language	31.2	49.4	36.2	22.1
Literature	31.0	45.1	26.5	10.4
Reading	36.7	60.1	20.6	30.0
Science	22.0	40.7	31.0	27.4
Social studies	31.0	51.2	29.8	19.3
Spelling	35.4	52.6	42.0	25.6

as a whole, as well as for boy-boy, girl-girl, and boy-girl combinations. The greatest amount of reduction in variation, in each measure, was found in the boy-boy pairs. The question arises: are brother-brother combinations more alike than the other groupings, or are the differences between unrelated boy-boy pairs merely greater? The means of the average differences indicated that the differences between unrelated boys were greater than those of the other groupings, and that the brother-brother pairs showed only slightly less variation than what was found for the total group.

The smallest amount of reduction in variation, 8.9 per cent, was for girl-girl combinations in arithmetic. Girls as a whole in this study were more similar to each other in arithmetic ability than they were in other achievement-age measures. The largest amount of reduction in variation, 60.4 per cent, was found for boy-boy combinations in mental age. Unrelated boys varied to a greater extent in mental ability than did the other groupings, and brothers were slightly less variable than was the total group.

In general, the content subjects showed less reduction in variation than did the skill subjects. Perhaps this was caused partly by the greater unreliability of tests in the content subjects. All the differences between the sib-pair means and unrelated pair means for the total group were significant at the five per cent level or lower, except for science. The number of sibling and unrelated pairs in science was the smallest of all the achievement measures, which may explain in part why the differences of the science means, though in the expected direction, did not meet the same test for significance.

TABLE III—CORRELATIONS FOR THE TOTAL GROUP OF SIBLING AND NONSIBLING PAIRS

Mental or Achievement Age	Siblings	Nonsiblings
Mental	0.71	0.27
Arithmetic	0.49	0.08
Education	0.59	0.02
Language	0.40	−0.09
Literature	0.41	−0.11
Reading	0.51	0.01
Science	0.39	−0.02
Social studies	0.64	0.14
Spelling	0.53	0.08

The correlation coefficients. The results secured by the method of correlation coefficients for the total group of sibling pairs and for the total group of unrelated pairs are given in Table III. For intelligence the correlation coefficient found for the siblings was high, indicating a marked relationship. The correlation for the nonrelated pairs was low, indicating a small relationship. Theoretically, the correlation expected for the unrelated pairs is zero. The fact that approximately ninety-five per cent of the children in this study have mental growth curves above the so-called norm may, in part, be responsible for the correlation for the unrelated pairs yielding a small relationship, rather than no relationship.

For achievement the correlation coefficients for the siblings were moderate, indicating a substantial relationship. The correlations found for the nonrelated pairs average +0.01 for the eight achievement measures.

According to the obtained t values, all the sibling correlations were significantly different from zero. It may be pointed out that these correlations substantiate each other and give some indication of the value of the correlations in the population, since they are all positive and are in a relatively narrow range.

The children involved in this study were a select and homogeneous group, since the large majority of them came from homes of superior economic, social, and intellectual status. It is possible that the differences between the unrelated pairs were smaller than they would have been if the total group had been less homogeneous.

CONCLUSIONS

All three methods of analysis utilized on the longitudinal growth records in this study, i.e., means of the average differences, percentage reduction of difference by family membership, and correlation, produced consistent results in describing the existence of a substantial amount of sibling resemblance in intelligence and achievement. Resemblances in intelligence were somewhat greater than they were in achievement.

REFERENCES

1) Willard C. Olson and Byron O. Hughes, *Manual for the Description of Growth in Age Units*. Ann Arbor, Edwards Letter Shop, 1950.

2) Sarah M. Schoonover, *Sibling Resemblances in Achievement*. Ann Arbor, University Microfilms, 1953.

BOOK REVIEWS

ARNOLD BERNSTEIN. *On the Nature of Psychotherapy*. New York: Doubleday & Company, pp. 36, 1954.

EMERY I. GONDOR. *Art and Play Therapy*. New York: Doubleday & Company, 1954, pp. 36.

ROBERT I. WATSON. *Psychology as a Profession*. New York: Doubleday & Company, 1954, pp. 65.

These are part of a series of Doubleday Papers in Psychology. They are paper-bound booklets which attempt to give concise statements about psychology in its various areas. They are written by individuals chosen for special interest and acquaintance with each special topic. They can be read by those who have not had a beginning course in psychology as well as by students of psychology who desire a special discussion of a particular subject.

Bernstein's booklet discusses, among other things, the nature of disease and of therapy, therapeutic acts, instruments and relationships, somatogenesis, psychogenesis, sociogenesis, the therapy group and the psychotherapeutic relationship, the patient's point of view, the point of view of the psychotherapist, the psychology of the therapist, and for further reading gives a well-selected list of books and papers.

Gondor's *Art and Play Therapy* deals in Section 1, with theory and practice; the child in therapy; and art and play in therapy. Section 2 deals with clinical applications: special technique for first contact; preliminary art and play therapy; and art and play therapy with retarded children. Section 3 presents three case histories illustrating first contacts with child and parents, achievement of a limited goal with a deeply disturbed child. There are two pages of selected references.

Watson's *Psychology as a Profession*, is of general interest and value for all who desire to know the problems of psychologists and their work in professional activities. Section 1 gives an orientation, relating to the nature of a profession, the professional concept in psychology and the kinds of specialization in psychology. Section 2 deals with the stabilizing professional factors in psychology: the American Psychological Association, training in professional psychology, ethics in the profession, and various organizations, national and local, that make for professional stabilization. Section 3 takes up unresolved intraprofessional issues: professional vs.

scientific influences, and practice on the subdoctoral level. Section 4 discusses various unresolved interprofessional issues, including the psychologist and psychotherapy, counseling, private practice, legal problems, and the relation of psychology to medicine. An excellent list of one hundred and eighteen references is appended.

These and others of the Doubleday Papers in Psychology will probably find extensive use for those who desire fairly nontechnical and generally good discussions of such subjects as The Natural Man, Nature and Nurture, Feelings and Emotions, Learning, Individual Development, etc.

A. S. Edwards

The University of Georgia

Nelson B. Henry, Editor. *Mental Health in Modern Education, Part 2, Yearbook of the National Society for the Study of Education.* Chicago: The University of Chicago Press, pp. 397, lxxiv.

Under the chairmanship of Paul A. Witty and a committee consisting of H. A. Carroll, P. T. Rankin, H. N. Rivlin and Ruth Strang, and twelve associated contributors, a valuable and for the most part well-condensed study is here presented on the problems of mental health in modern education.

Although rather long delayed as a major subject of examination in this series of studies, it is very timely. It is one of several recent publications on the subject which many believe to be more than half of the whole problem of health.

The importance of mental health is emphasized in a number of places in the volume. One such emphasis is expressed by Shibler (pp. 303–4) who writes: "The Citizenship Education Study in Detroit points to one conclusive fact: The basis for training for citizenship in the schools is good mental health."

This comprehensive study deals with many phases of the subject under the following large heads: the needs for the program and the rôle of this yearbook in the improvement of mental health; the history and present status of the mental health movement; conditions affecting mental health in the classroom; problems and practices related to mental health at different educational levels; personal and professional development of the teacher; and mental health for today and tomorrow.

In more detailed analysis we find such subjects as motivation and learning in mental health; the rôle of the home, of the school, and of the community in the mental health program; special problems at the preschool level, in the primary grades, in the intermediate grades, in the high school, at the college level, and an evaluation of present programs.

Some brief discussion is given of the contributions of anthropology, sociology, psychology, general semantics, philosophy, religion and law. Reference is made to mental mechanisms, psychoneuroses, emotional security, emotional conflict, counseling, causes of mental and emotional deviation, adjustments, neurotic and psychotic types.

Many but not all of the chapters have excellent lists of references at the ends; a large number of footnotes accompany some of the chapters.

The great comprehensiveness of the study makes impossible the more detailed discussions which many will want, but which they can find in the special articles and books on the various subjects. There is an index of six pages which is well made but not quite as complete as it might well be.

The appendices (pp. i–lxxiv) present the constitution and by-laws; minutes of the Atlantic city meeting; synopses of some proceedings of the Board of Directors; report of the treasurer; a list of members; information concerning the society; and a list of the publications of the society.

The book is a very valuable contribution to a most important subject—one of the most important for parents and teachers. It gives evidence that the lives, efforts and hopes of such men as Clifford Beers and W. H. Burnham were not in vain.

A. S. Edwards

The University of Georgia

PUBLICATIONS RECEIVED

VERGILIUS FERM AND OTHERS. *A Dictionary of Pastoral Psychology*. New York: Philosophical Library, 1955, pp. 336. $6.00.

Fifty-Fifth Annual Report of the Superintendent of Schools City of New York: Statistical Section, School Year 1952-1953. Prepared by the Bureau of Administrative and Budgetary Research. Board of Education, New York City.

Fifty-Sixth Annual Report of the Superintendent of Schools of New York City. Prepared by the Bureau of Administrative and Budgetary Research. Board of Education. New York City.

ARTHUR E. FINK, EVERETT E. WILSON AND MERRILL B. CONOVER. *The Field of Social Work*. Third Edition. New York: Henry Holt & Co., 1955, pp. 630. $5.25.

ARTHUR W. FOSHAY, KENNETH D. WANN, AND ASSOCIATES. *Children's Social Values: An Action Research Study*. New York: Bureau of Publications, Teachers College, Columbia University, 1954, pp. 323. $3.50.

LAWRENCE K. FRANK. *Individual Development*. New York: Doubleday & Co., Inc., 1955, pp. 52 (paper). $0.85.

D. C. FRASER. *A Psychological Glossary*. Cambridge: W. Heffer & Sons Ltd., 1954, pp. 40 (paper).

FRANK S. FREEMAN. *Theory and Practice of Psychological Testing*. Revised Edition. New York: Henry Holt & Co., 1955, pp. 609. $5.25.

ELIOT FREIDSON, EDITOR. *Student Government, Student Leaders and the American College*. Philadelphia, Pa.: United States National Student Association, 1955, pp. 88 (paper).

KARL C. GARRISON AND J. STANLEY GRAY. *Educational Psychology*. New York: Appleton-Century-Crofts, Inc., 1955, pp. 505. $5.00.

FRANK H. GERI. *Illustrated Games and Rhythms for Children. Primary Grades*. New York: Prentice-Hall, Inc., 1955, pp. 196. $2.95.

EDWIN E. GHISELLI AND CLARENCE W. BROWN. *Personnel and Industrial Psychology*. Second Edition. New York: McGraw-Hill Book Co., 1955, pp. 491. $6.00.

H. HARRY GILES. *Human Dynamics and Human Relations Education*. New York: New York University Press, 1954, pp. 108 (paper).

JAMES M. GILLESPIE AND GORDON W. ALLPORT. *Youth's Outlook*

on the Future. Garden City, New York: Doubleday & Company, Inc., 1955, pp. 61 (paper). $0.85.

EMERY I. GONDOR. *Art and Play Therapy.* Garden City, New York: Doubleday & Company, Inc., 1954, pp. 61 (paper). $0.95.

Graphic Arts Shop, Teacher's Work Manual. Board of Education of the City of New York. 1954, pp. 77 (paper).

HARRY A. GREENE AND HELEN I. STAPP. *Greene-Stapp Language Abilities Test.* Yonkers-on-Hudson, N. Y.: World Book Company (paper).

JULIEN GREENLEE. *Better Teaching Through Elementary Science.* Dubuque, Iowa: Wm. C. Brown Company, 1954, pp. 204 (paper). $2.50.

JULIAN GREENLEE. *Teaching Science to Children.* Dubuque, Iowa: Wm. C. Brown Company, 1955, pp. 185 (paper).

WILLIAM W. GRINGS. *Laboratory Instrumentation in Psychology.* Palo Alto, California: The National Press, 1954, pp. 282. $4.75.

J. E. GRINNELL, RAYMOND J. YOUNG AND OTHERS. *The School and the Community: Educational and Public Relations.* New York: The Ronald Press Co., 1955, pp. 444. $5.50.

SIDONIE MATSNER GRUENBERG, Editor. *The Encyclopedia of Child Care and Guidance.* New York: Doubleday & Co., Inc., 1954, pp. 1016. $7.50.

A Guide to Some Educational Resources in the City of New York. Curriculum Bulletin, 1953-1954 Series, Number 7. Board of Education of New York City, 1954, pp. 107 (paper).

J. P. GUILFORD. *Psychometric Methods.* Second Edition. New York: McGraw-Hill Book Co., 1954, pp. 507. $8.50.

A. PAUL HARE, EDGAR F. BORGATTA AND ROBERT F. BALES. *Small Groups.* New York: Alfred A. Knopf, 1955, pp. 666.

THISTLE Y. HARRIS. *The Teaching of Nature Study.* Melbourne: Australian Council for Educational Research, 1954, pp. 100 (paper).

Health Supervision of Young Children. A Guide for Practicing Physicians and Child Health Conference Personnel. New York: American Public Health Association, 1955, pp. 179. (paper) $2.00.

NELSON B. HENRY, Editor. *Mental Health in Modern Education. The Fifty-fourth Yearbook of the National Society for the Study of Education.* Part II. Chicago: The University of Chicago Press, 1955, pp. 397 (paper).

VIRGIL E. HERRICK AND LELAND B. JACOBS. *Children and the Lan-*

guage Arts. Englewood Cliffs, N. J.: Prentice-Hall Inc., 1955, pp. 524. $5.50.

Home Economics, Clothing and Household Fabrics, Grades 7, 8 and 9. Board of Education of New York City, 1954, pp. 120 (paper).

THOMAS W. HOPKINS, HARRY V. BICE AND KATHRYN C. COLTON. *Evaluation and Education of the Cerebral Palsied Child.* Washington, D. C.: International Council for Exceptional Children, 1954 pp. 114 (paper).

JAMES L. HYMES, JR. *Behavior and Misbehavior.* New York: Prentice-Hall, Inc., 1955, pp. 140.

JAMES L. HYMES, JR. *A Child Development Point of View.* New York: Prentice-Hall, Inc., 1955, pp. 145.

ARTHUR T. JERSILD. When Teachers Face Themselves. Columbia University: Teachers College, Bureau of Publications, 1955, pp. 169. $3.65.

WENDELL JOHNSON AND RALPH R. LEUTENEGGER. *Stuttering in Children and Adults.* Minneapolis: University of Minnesota Press, 1955, pp. 472. $5.00.

HARRY W. KARN AND JOSEPH WEITZ. *An Introduction to Psychology.* New York: John Wiley & Sons, Inc., 1955, pp. 315. $3.90.

S. DONALD KARL AND BARBARA L. DIEHL. Editors. *The College Handbook.* New York: College Entrance Examination Board, 1955, pp. 399 (paper).

FRED S. KELLER. *Learning Reinforcement Theory.* Garden City, New York: Doubleday and Company, Inc., 1954, pp. 37 (paper). $0.85.

VICTOR H. KELLEY AND HARRY A. GREENE. *Kelley-Greene Reading Comprehension Test.* Yonkers-on-Hudson, N. Y.: World Book Company (paper).

ED KILMAN AND THEON WRIGHT. *Hugh Roy Cullen: A Story of American Opportunity.* New York: Prentice-Hall, Inc., 1954, pp. 376. $4.00.

LEONARD KORNBERG. *A Class for Disturbed Children: A Case Study and Its Meaning for Education.* New York: Bureau of Publications, Teachers College, Columbia University, 1955, pp. 157. $3.75.

RUTH KOTINSKY AND HELEN L. WITMER, Editors. *Community Programs for Mental Health.* Cambridge, Mass., 1955, pp. 362. $5.00.

W. D. LAMONT. *The Value Judgement.* New York: Philosophical Library, 1955, pp. 335. $6.00.